2

AD-A269 860

July 12, 1993 Final Report 01/25/91 to 09/39/93

System Dynamics and Computer Modeling Expertise: A Conceptual Structure Research Inquiry.

GRANT # N00014-91-J-1360

MARK E. SIEGEL

University of the District of Columbia
4200 Connecticut Avenue, N. W.
Washington, D. C. 20008

Cognitive Science Program
Office of Naval Research, Department of the Navy
800 N. Quincy Street
Arlington, VA 22217-5660

None

Approved for public release; Distribution unlimited.

This research examined the conceptual structures of 22 expert system dynamists from 10 countries in order to investigate the extent to which representations of their knowledge revealed similarities in conceptual structure properties. The results revealedless similarity than has been found in experts from other fields. It was hypothesized that the hetrogenity of the sample combined with the fact that the filed itself is still developing were the major contributor factors that produced the results. The methodology was then employed with graduate students to examine shifts in their conceptual structures as they acquired system dynamics knowledge. The results were reported in an appendix to the main study.

Conceptual structures; expertise; system dynamics; computer modeling and simularion. 62

Unclassified Unclassified Unclassified SAR

AD-A269 860

②

July 12, 1993 Final Report 01/25/91 to 09/39/93

System Dynamics and Computer Modeling Expertise: A Conceptual Structure Research Inquiry.

GRANT # N00014-91-J-1360

MARK E. SIEGEL

University of the District of Columbia
4200 Connecticut Avenue, N. W.
Washington, D. C. 20008

Cognitive Science Program
Office of Naval Research, Department of the Navy
800 N. Quincy Street
Arlington, VA 22217-5660

None

Approved for public release; Distribution unlimited.

This research examined the conceptual structures of 22 expert system dynamists from 10 countries in order to investigate the extent to which representations of their knowledge revealed similarities in conceptual structure properties. The results revealedless similarity than has been found in experts from other fields. It was hypothesized that the hetrogenity of the sample combined with the fact that the filed itself is still developing were the major contributor factors that produced the results. The methodology was then employed with graduate students to examine shifts in their conceptual structures as they acquired system dynamics knowledge. The results were reported in an appendix to the main study.

Conceptual structures; expertise; system dynamics; computer modeling and simularion.

62

Unclassified Unclassified Unclassified SAR

12

CONCEPTUAL STRUCTURES OF EXPERT SYSTEM DYNAMISTS[1]

Mark Everet Siegel[2]
University of the District of Columbia

INTRODUCTION

Learning to be a scientist entails the acquisition and use of different paradigms or ways of understanding that are counterintuitive to common experience. This viewpoint applied as a pedagogical principle rests on several assertions. First, it is essential for science students to develop causal reasoning skills to understand complex dynamic problems. Second, students require an organizing framework which

[1]This research was sponsored by the Cognitive Science Program, Office of Naval Research, under Grant No. N00014-91-J-1360. Approved for public release; distribution unlimited.

[2] The author wishes to acknowledge the advice and assistance of Drs.Tim Goldsmith and Roger Schvaneveldt for their help with graph theory and the Pathfinder software; Dr. Richard Colker for programing of the PathPrep software tools; Drs. Steve Peterson and George Richardson as expert advisors on System Dynamics theory and concept selection; and Dr. Susan Chipman for her encouragement and support.

emphasizes analytic, analogical, and synthetic cognitive processing for achievement in advanced knowledge acquisition. Third, the ability to represent dynamic causal problems as computational models and to run exploratory simulation experiments with the models, is a crucial step toward achieving an in-depth understanding of causal problems. Forth, an interactive learning environment is most suitable for learningcomputer modeling/simulation skills. Fifth, the knowledge domain associated with generative computer modeling/simulation has structural properties with distinctive patterns indicative of expertise level. Sixth, shifts in novice conceptual structures can be used to diagnose knowledge acquisition and guide the presentation of subsequent instructional messages.

Accepting the validity of assertions one through four, the goal of this research is to investigate empirically the fifth claim, and in particular the similarity of conceptual structures of expert system dynamists. This is necessary in order to provide what Goldsmith and Teague (1993) and Johnson, Acton, and Goldsmith (1993) mean by a "referent -based" evaluation standard on which to judge students' knowledge acquisition.

The report begins with a discussion of paradigm shifts, and then outlines in some detail a theory of advanced knowledge acquisition which this author believe explains the difficulty students experience

learning to construct their own models from text descriptions. Next, there is a review of conceptual structure methodology and issues useful in deciding graph/network similarity. The research activity is then described followed by an analysis and summary section. Included in the appendix are examples of conceptual structures and Stella II models. The author acknowledges his debt to those whose work has been influential, and will report extensively on that work to make clear the basis for the present basic research effort.

Paradigm Shifts:

A necessary condition for the achievement of deep understanding is prior knowledge that serves as a foundation for new learning (Bransford, 1979). In addition, there is a widely held view that schemata comprise the "basic units" of knowledge structures. Rumelhart and Norman (1981a) extend this idea by listing three ways in which existing schemata are modified by experience. Accretion refers to a resultant change in knowledge structure through the gradual accumulation of facts within existing schemata. Tuning pertains to changes in the categories used to interpret information so that the accuracy of the schema is in consort with the data. Restructuring applies to changes in the knowledge that involves the creation of new structures, whose purpose is to reinterpret prior information or to explain new information. According to Vosniadou and Brewer (1987), "Restructuring represents

the most radical form of knowledge acquisition within a prior knowledge framework."

While Piaget (1929) suggested a global restructuring process to account for developmental changes, many contemporary researchers contend that restructuring should be considered in terms of domain specific experiences rather than in terms of some general !ogic capability (Carey, 1985b). Vosniadou and Brewer note that " . . . the process of conceptual change may involve not the simple modification of an existing underdeveloped conceptual structure, but rather the formation of radically new conceptual models, something that Wittrock (1981) and Osborne and Wittrock (1983) called 'generative learning'.".

The literature generally characterizes domain specific restructuring into two types (Carey, 1985a). On the one hand, there is weak restructuring which emphasizes the accrual of new facts and the establishment of new relationships between existing ideas (Chi, Glaser, and Ross, 1982; Voss, Green, Post, and Penner, 1983). On the other hand, radical restructuring involves changes to the central concepts and the relationships between them (see diSessa (1982; McCloskey, 1983; Wiser and Carey, 1983). Much of this work has been conducted with the goal of understanding novice/expert differences. In essence these studies suggest that being a novice is not the simple matter of having an

insufficiency of information, but a basically different theory about the knowledge domain in terms of both domain structure and individual concepts.

Yet every knowledge domain has data for which its theories cannot consistently account. At some point theory modification is recognized as no longer a viable approach to achieving a satisfactory explanation and a shift to a new paradigm follows.

Thus, both the weak and the radical restructuring play a role in knowledge acquisition and understanding. However, as Vosniadou and Brewer state, " . . . only the radical type can account for the emergence of completely new theories or new paradigms out of existing structures" (p55).

The question now arises about what processes could be employed to engender radical restructuring? Several possibilities have been suggested including differentiation and coalescence (Carey, 1985b); Socratic dialogue (Champagne, Klopfer, and Gunstone, 1982): explanation (DeJong and Mooney, 1986); generalization and specialization (Rumelhart and Ortony, 1977); metaphors, physical models, and analogies (Vosniadou and Schommer, 1986); and schema induction (Rumblhart and Norman, 1981b).

For both children and adults analogies can have different roles during restructuring: spontaneous restructuring of new schemata and explicit teaching of new schemata. Gick and Holyoak (1980, 1983) have reported that the spontaneous generation of analogy is difficult to achieve in an experimental situation. But when used as an explicit instructional strategy for constructing new schemata or restructuring schemata, analogy can be very effective (see also Gentner, 1981; Vosniadou and Ortony, 1983; Vosniadou and Schommer, 1986).

Advanced Knowledge Acquisition and Misconceptions.

Too often students use "common sense" or other ill defined cognitive strategies when reasoning about complex causal problems, frequently leading to pervasive and pernicious misconceptions.

Feltovich, Spiro, and Coulson (1988) provide a theory of advanced knowledge acquisition that explains the process underlying the development of conceptual structures. The state, "It is with regard to advanced knowledge acquisition for complex material that current theories of learning are most deficient and current educational methods least effective . . . little is known about the acquisition of advanced understandings found in expertise or about the best educational methods for fostering them.".

Feltovich, Spiro, and Coulson argue that the errors in understanding are most likely not to result from a single source. The Multiplicity factor refers to that complex of influences ranging from those associated with the learner, the educational process, and even the research literature itself. The Interdependency factor refers to misconceptions as reciprocating networks of faulty component ideas that are mutually reinforcing and therefore support the general misconception. The Oversimplification factor pertains to complex knowledge being aggregated to the extent that the essential information is unavailable. Thus, an extensive analysis of the knowledge domain is a necessary condition for understanding students' problems in achieving deep understanding, and in being able to rectify such problems.

Turning to the matter of why concepts are "difficult," Feltovitch, Spiro, and Coulson believe that unusual demands are made on the person's cognitive system. These demands may be categorized four ways: 1) unusual demands on working memory; 2) unusual demands on formal representation"; 3) unusual demands on "intuition" or prior knowledge; and 4) unusual demands on notions of regularity. There emerges in this theory the important notion of reciprocation among families of related ideas which may explain difficulties students have learning complex phenomena.

Feltovitch, Spiro, and Coulson argue that the more typical instructional practices do not serve the student well because complexity is often dealt with by transforming it into a simpler form with the expectation that subsequent instruction will incrementally add greater sophistication (Glaser, 1984 is cited as an example of this "scaffold" model). This "simplification approach" is criticized on two accounts. First, based on research findings that initially simplified instructional approaches tend often to impede the later acquisition of complexity (Feltovitch, Coulson, and Spiro, 1986).

The second objection to the "scaffold" model is that early experiences with complexity reduction lead to simplification of more advanced knowledge. In other words, oversimplification becomes the operating rule. Feltovitch, Spiro, and Coulson thus distinguish between simpler and advanced knowledge acquisition places the highest priority on getting the ideas right (their emphasis). Furthermore, the two approaches differ in terms of the evidence needed to assess understanding. The scaffold models emphasize imitative memory for key terms and definitions. With advanced knowledge acquisition the emphases are on the novel "use, application, or transfer of explicitly taught material." In fact, they claim, "In fact, it seems likely that those factors that promote accurate memory, (e.g., tightly compartmentalized,

insular mental representations are antithetical (their emphasis) to the development of usable/applicable knowledge." In this framework, understanding is conceived as a process by which what is learned at one time influences subsequent learning about that idea at another time.

The expert-novice literature, to which I now turn, reveals important differences between these two groups vis-a-vis knowledge. As summarized by Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tuchker, and DeMaio (1985) expert cognitive characteristics include superior performance on recall tasks in which meaningful material is used (although there are no differences when the same material is recalled in random order), increases in experience leads to higher intragroup agreement among experts in relation to memory structure and organization, and experts have larger chunk sizes and more chunks. This and other evidence indicate that there are major cognitive changes that occur during the transition from novice to expert. This author believe that such changes manifest themselves through alterations in the structural properties of concept networks and can be measured using the methodology discussed below.

System Dynamics and Computer Modeling.

Until the arrival of powerful personal computers, modeling and simulation required expert programming knowledge and access to

expensive hardware resources. Undergraduate students and faculty outside computer science departments had almost no experience with modeling as a pedagogical technology.

Yet computer modeling is more than a technique. Its purpose is to improve understanding through the study of interrelationships among a problem's parts in terms that require every relevant assumption to be specified operationally. Models are not truth. They are paths to insight. This is their value in education.

The basic premise of Systems Dynamics (the systems theory adopted here) is that phenomena which change over time can be understood by focusing on the structures that find the relationships among the relevant components, rather than on discrete instances associated with their functioning (i.e. localized symptoms). Phenomena are composed of components which have specific functions, and when viewed collectively, have some common goal or purpose. By employing formalisms which enable the student to represent his/her problem of study in increasingly abstract terms, students acquire a variety of cognitive skills important to the development of causal reasoning and understanding. Students learn how to make explicit their mental models of phenomena using a vocabulary through which both qualitative and quantitative ideas can be expressed. This leads to effective

communication in both instructive and collaborative learning experiences.

Forrester (1976), who set forth the initial theoretical framework of modern System Dynamics, wrote a programmed text introduction to the field. Roberts, Anderson, Deal, Garet, and Shaffer (1983) wrote the first introductory level text (with an extensive workbook) to teach system dynamics using a microcomputer (Apple™ II). One of the most widely used special purpose mainframe modeling languages available was Dynamo™ and its contemporary versions (for both large and small computers) remains a favored choice by many people. Richmond and Peterson (1985) developed the modeling language, STELLA™ (Structural Thinking Experimental Learning Laboratory with Animation) which takes advantage of many user oriented features such as icons, interactive graphics, and animation of the Macintosh™ computer. With STELLA™ students are not confined to writing equations as the only means of model construction. Rather, they utilize menu tools and icons to represent the system structure. The basic forms of the needed equations are generated concurrently by STELLA™. (Examples of Stella™ features are provided in the Appendix).

One series of studies on teaching systems dynamics and computer modeling to improve causal reasoning has been conducted by

Mandinach (1989, 1988a, 1988b, 1987). The Systems Training and Curriculum Innovation (STACI) Project looked at both domain knowledge acquisition and problem solving skills development in high school students who received systems dynamics and Stella™ instruction integrated into their curricula. Bear in mind that this series of research was not conducted in the context of an interactive learning laboratory. Five subject areas were selected: (general physical science, biology, physics, chemistry, and a special history course). With the exception of the latter, there were "systems" and "traditional" classes for each topic. Several test instruments were used to assess students' content and problem solving knowledge.It is not the purpose here to review, in depth, all of the aspects of this research. There are, however, implications relevant to the purposes of this proposal. While the data did not provide conclusive evidence that major cognitive restructuring occurred for either content acquisition or problem solving skills, the authors suggest that a longer term research effort may be needed to show such results. Nevertheless, they conclude, "There were, however, qualitative differences in the way systems and traditional students approached problems" (pp 32). For example, the systems group focused on the components of a system's behavior that changed over time whereas the traditionally taught students concentrated on terminology. Another very important result to emerge from these studies is the recognition that the development of systems thinking occurs slowly, that

explic!t connections between domain content and systems theory must be made; that students had difficulty modeling problems even with a basic understanding of systems theory; and that teachers selectively emphasized various aspects of systems theory resulting in students from different classes having different degrees of systems theory understanding. As will is noted below, the achievement of expert level knowledge requires a great deal of practice over extend time periods. In discussions with some members of the research team, and after a review of their publications, it is the interpretation here that the crucial variable in determining student success is the knowledge and commitment of the teacher to integrate systems theory and modeling into the course. The classroom teacher can achieve what the interactive system cannot achieve, meeting specific individual needs.

GRAPH THEORY AND CONCEPTUAL STRUCTURES:

A useful metaphor for a domain of knowledge is as a related collection of concepts represented by a node-link graph. When a metric is applied to a graph's links the graph becomes a network and mathematical transformations can be employed which extract structural information (Chartrand,1977). Schvaneveldt (1990) presents both a theoretical framework and supporting research on the use of graph/network analyses to the assessment of knowledge structure. This work should be consulted by the interested reader particularly regarding

the Pathfinder™ conceptual structure analysis software which will play a crucial role in this project.

There is no absolute standard for comparing network structure representations. With regard to Pathfinder networks, two different approaches have been used with the goal of applying rules or functions to obtain a real number indicative of similarity. A real number that indicates multigraph similarity may be useful in assessing the status of student' knowledge vis-a-vis the instructor, possible sources of misconceptions, or as a guide to the next instructional message that will be provided.

Each graph has a particular arrangement of its node-link components. One view of structure starts with the idea of primitive relations. The more complex node-link patterns are combinations of the primitive relations and constitute higher order relations. For example, in system dynamics theory, the primitive relation "flow" (or rate) is a basic building block for the higher order relation " compounding process". The higher the order of a relation the more it provides information about a particular structure. In this framework structure is an emergent property.

The alternative approach Goldsmith and Davenport discuss is to view structure as an entity, a collection of sub-objects. Components P and Q together form the larger component R.

The entity versus emergent property viewpoints determines the methodology, the evidence collected, and the conclusions reached regarding structural similarity. In the entity case, one might partition the edges until a common subgraph is achieved (Ulam's Method(see Graham, 1987)). This method assumes that the graphs' nodes are unlabeled, which is usually not the case for knowledge domains.

The alternative, which Goldsmith and Davenport adopt, is to use two higher order property relations (distance and neighborhood) as the basis for similarity assessment. The neighborhood for any given node is those nodes within a path length of one unit of the target node, excluding the target itself.

One analysis strategy employed by compute the correlation coefficient of two graphs distance vectors[3](An example for Pathfinder networks can be found in Schvaneveldt, Durso, Goldsmith, Breen, Cooke, Tucker, and DeMaio,1985).

The nodes which form a neighborhood make up a set. To achieve a ratio of shared elements to total elements between the same

---

[3] Adjacency and distance matrices of undirected graphs are composed of two triangles that mirror each other and share a common hypotenuse of zero cell entries indicative of judging a node with itself. Thus the vector of n nodes will have $(n^2-n)/2$ distances.

neighborhoods of different graphs, intersection, union, and cardinality are determined.

Goldsmith and Davenport offer the following rule for assessing graph similarity. They state, " . . . an index of similarity for a common node in two graphs is the cardinality of the intersection of the node neighborhoods divided by the cardinality of the union of the neighborhoods. One measure of overall graph similarity is the mean of these n values. This measure will vary from zero to one with higher values indicating greater similarity(p78).

There is no "best" approach for determining similarity. Path distance involves node pairs and graphs are compared using this unit. On the other hand, neighborhood analysis is based only on nodes of the same unit distance, and a real number index for inter-graph comparisons is based on set theory. Thus, the specific hypothesis determines the methodology to be employed. If the question about the knowledge domain centers on the psychological distance between concepts, path analysis is the method of choice. On the other hand, if the question concerns the relationships among concepts in terms of the configural structure of the knowledge domain, the neighborhood method seems better suited.

Current Funded Research

M.S., who will be the Principal Investigator is at present the Principal Investigator of a sponsored research project funded by the Cognitive Sciences Program of the Office of Naval Research. That project will be completed no later than September 30, 1993. Its purpose was to investigate the similarities of the cognitive structures of expert system dynamists. A software tool was developed (PathPrep) which aids experts of any field to select subsets of "core or central" concepts from a larger set that comprises a body of knowledge. PathPrep then uses that information to solicit pairwise judgment data from subjects. The output is placed in a file used by Pathfinder to analyze the data. Concepts terms provided in the Appendix were derived by system dynamists advising the ONR project. Twenty two expert subjects in ten countries were sent PathPrep disks and returned similarity data. These results were processed by Pathfinder and are at present in the analysis stage. The PathPrep/Pathfinder combination has proven its usefulness in cognitive structure research.

The data analysis yielded similarity scores (.31) lower than the usual range obtained from experts of other fields. A consultant (Goldsmith) has suggested that the data reflect the heterogeneity of the sample (Ss were from ten countries in North America, Europe, and Asia). In virtually every other study of this type, the subjects were from the

same departments, schools, or cultures. This author know of no other research in system dynamics that is either cross-cultural vis-a-vis experts or comparing novices with experts in terms of building computer models. Nor is anything known about conceptual structure shifts and interactive learning environments.

RESULTS

The data were first examined with a frequency distribution of subject by scale response choice to determine the pattern of subjects use of the 9 point values available to them. As with the subsequent analyses, the following code scheme was employed.

TABLE #1
Code Scheme

#	Concept Terms
1	ACCUMULATION
2	AGGREGATION
3	BOUNDARY
4	CAUSATION
5	COMPENSATING FEEDBACK
6	CONTROL STRUCTURE
7	DOMINANCE SHIFTS
8	DYNAMIC HYPOTHESIS
9	DYNAMIC THINKING
10	ENDOGENOUS VIEW
11	EXPONENTIAL DECAY
12	EXPONENTIAL GROWTH
13	EXTREME CONDITIONS TEST
14	LEVEL (STOCK) VARIABLE
15	NEGATIVE FEEDBACK
16	NONLINEAR RELATIONSHIP
17	OPERATIONAL THINKING
18	OSCILLATION
19	OVERSHOOT AND COLLAPSE
20	PARAMETER INSENSITIVITY
21	POSITIVE FEEDBACK
22	RATE (FLOW) VARIABLE
23	ROBUSTNESS
24	SENSITIVITY TESTING
25	SIGMOIDAL GROWTH

#	Subject
1	CA09
2	CA20
3	GR29
4	IN28
5	IS11
6	NE26
7	NE27
8	NO15
9	NO30
10	TI19
11	TR04Y
12	UK05
13	UK24
14	UK25
15	US07
16	US08
17	US10
18	US14
19	US17
20	US18
21	US21
22	US23

The first task was to examine the frequency distribution of each subject's 300 ratings over the nine point rating scale. Table 2 shows the frequency distribution of the raw ratings values for each subject along with an overall frequency distribution across all subjects and all pairs.

Table 2. Frequency distribution of the raw ratings values by subject.

	Similarity Rating Values								
Subject	1	2	3	4	5	6	7	8	9
1	3	30	26	2	32	20	51	120	16
2	14	24	14	5	6	39	62	72	64
3	2	8	24	56	108	35	35	15	17
4	70	67	15	4	33	20	26	43	22
5	11	3	29	35	47	65	65	32	13
6	11	30	38	25	56	56	57	19	8
7	0	75	58	42	30	25	61	9	0
8	21	31	25	28	61	48	51	23	12
9	15	14	22	17	25	53	61	69	24
10	10	47	29	3	9	56	56	46	44
11	55	87	27	29	23	23	21	19	16
12	26	56	10	4	78	7	44	47	28
13	25	53	33	15	44	22	27	40	41
14	81	52	38	17	26	33	28	16	9
15	130	76	35	17	15	16	6	3	2
16	3	35	32	15	76	24	45	28	42
17	5	60	55	30	32	30	25	32	31
18	0	6	28	73	69	57	39	23	5
18	7	28	50	20	16	40	81	42	16
20	8	8	45	53	20	36	54	28	48
21	110	40	17	6	9	26	43	36	13
22	16	34	36	25	14	46	48	17	64
Sum	623	864	686	521	829	777	986	779	535
Frequency	0.094	0.131	0.104	0.079	0.126	0.118	0.149	0.118	0.081

With the exception of subjects 7 and 18, the remaining 20 people used all nine values of the rating scale. The general form of the distribution tended to be normally distributed or rectilinear, although subject 15's distribution was highly skewed with proportionately many more pairs assigned low similarity values.

Scope:

The scope of a concept is defined by the number of other concepts that are seen to be related to it. High scope terms are generally category labels or other abstract terms. High scope concepts also have many connections in Pathfinder networks.

The scope of a concept is operationally defined by the number of other concepts seen to be related to it. To determine the number of related concepts from a multi-value scale (e.g., 1 to 9), it is necessary to define a rating score to serve as a cutoff mark whose purpose is to define operationally the related and unrelated scale values. Rather than selecting a single value and appiying

it to all subjects, a score for each individual subject was selected based on his/her cumulative frequency distribution of rating values. This method allows for individual differences in subjects' use of the scale. The score closest to the 25th percentile of relatedness was selected as the cutoff value. Similarity values equal to or greater than this score are

presumed to be related and values below this cutoff score are presumed to be unrelated.

The cutoff value was then used to transform the similarity values for every subject. Each subject's set of 1 to 9 similarity values was transformed into a set of 0's (unrelated) and 1's (related). From these 0/1 matrices, the number of subjects who rated each pair as related was counted. Table 3 shows the results of this analysis. It is clear from Table 3 that some of the terms were seen to be related to many other concepts, whereas other terms had fewer associates.

Table 3. The number of subjects (out of 22) who rated Term 1 as related to Term 2, where related was defined by each subject's 25th percentile score.

Term 2

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
1	0	4	1	1	1	3	3	5	10	3	3	7	0	20	0	0	1	2	2	0	7	10	1	0	3
2	4	0	14	1	0	0	0	7	5	1	0	0	0	6	0	1	2	0	2	3	1	2	4	5	0
3	1	14	0	7	4	5	2	12	10	18	1	1	12	2	0	5	3	1	0	4	1	1	9	7	1
4	1	1	7	0	9	13	6	18	18	13	2	4	2	1	10	4	8	2	4	3	6	4	4	3	4
5	1	0	4	9	0	14	6	9	15	13	3	0	2	1	16	3	5	7	3	17	1	0	13	10	8
6	3	0	5	13	14	0	9	9	12	10	1	2	2	5	16	2	4	8	7	6	5	7	7	5	5
7	3	0	2	6	6	9	0	13	13	13	1	8	8	0	2	17	2	6	15	5	6	0	8	9	18
8	5	7	12	18	9	9	13	0	20	17	3	7	3	3	8	7	6	4	8	3	8	2	5	2	8
9	10	5	10	18	15	12	13	20	0	19	7	6	6	5	14	6	7	8	10	8	7	8	6	4	8
T 10	3	1	18	13	13	10	13	17	19	0	3	3	4	3	9	1	5	5	4	6	6	4	10	2	3
e 11	3	0	1	2	3	1	1	3	7	3	0	13	2	0	14	6	0	0	2	2	10	1	0	0	2
r 12	7	0	1	4	0	2	8	7	6	3	13	0	3	2	1	2	1	1	10	1	21	1	0	1	10
m 13	0	0	12	2	2	2	8	3	6	4	2	3	0	0	2	9	2	1	3	13	1	1	16	20	1
14	20	6	2	1	1	5	0	3	5	3	0	2	0	0	1	1	2	4	1	0	3	17	1	0	0
1 15	0	0	0	10	16	16	2	8	14	9	14	1	2	1	0	2	3	15	10	12	9	1	11	2	10
16	0	1	5	4	3	2	17	7	6	1	6	2	9	1	2	0	3	6	16	10	3	1	6	4	15
17	1	2	3	8	5	4	2	6	7	5	0	1	2	2	3	3	0	1	1	1	2	2	2	1	0
18	2	0	1	2	7	8	6	4	8	5	0	1	1	4	15	6	1	0	6	1	1	1	0	1	0
18	2	2	0	4	3	7	15	8	10	4	2	10	3	1	10	16	1	6	0	1	9	0	0	0	6
20	0	3	4	3	17	6	5	3	8	6	2	1	13	0	12	10	1	1	1	0	2	0	19	20	2
21	7	1	1	6	1	5	6	8	7	6	10	21	1	3	9	3	2	1	9	2	0	0	1	0	6
22	10	2	1	4	0	7	0	2	8	4	1	1	1	17	1	1	2	1	0	0	0	0	0	0	3
23	1	4	9	4	13	7	8	5	6	10	0	0	16	1	11	6	2	0	0	19	1	0	0	21	4
24	0	5	7	3	10	5	9	2	4	2	0	1	20	0	2	4	1	1	0	20	0	0	21	0	1
25	3	0	1	4	8	5	18	8	8	3	2	10	1	0	10	15	0	0	6	2	6	3	4	1	0

Finally, to obtain an index of each concept's scope the mean number of associates for each term was determined by averaging across the values in each row. Table 4 shows these scope values along with terms ranked from highest to lowest scope. Thus, DYNAMIC THINKING is seen to be the most general term and AGGREGATION the most specific term.

Table 3. Scope values for the 25 terms ranked from highest scope to lowest

scope.

#	Term	Scope
9	DYNAMIC THINKING	9.67
8	DYNAMIC HYPOTHESIS	7.79
10	ENDOGENOUS VIEW	7.29
7	DOMINANCE SHIFTS	7.08
15	NEGATIVE FEEDBACK	7.00
5	COMPENSATING FEEDBACK	6.67
6	CONTROL STRUCTURE	6.54
23	ROBUSTNESS	6.17
4	CAUSATION	6.13
20	PARAMETER INSENSITIVITY	5.79
16	NONLINEAR RELATIONSHIP	5.42
3	BOUNDARY	5.04
19	OVERSHOOT AND COLLAPSE	5.00
24	SENSITIVITY TESTING	4.92
25	SIGMOIDAL GROWTH	4.92
21	POSITIVE FEEDBACK	4.83
13	EXTREME CONDITIONS TEST	4.71
12	EXPONENTIAL GROWTH	4.38
1	ACCUMULATION	3.63
18	OSCILLATION	3.38
14	LEVEL (STOCK) VARIABLE	3.25
11	EXPONENTIAL DECAY	3.17
22	RATE (FLOW) VARIABLE	2.75
17	OPERATIONAL THINKING	2.67
2	AGGREGATION	2.42

Coherence

The coherence of each subject's proximity file was next assessed. Coherence is a measure of the internal consistency or internal connectedness of a set of pairwise relations. The basic idea is that if two concepts are seen to be related, then how they in turn relate to still other concepts should be similar. To the extent that such agreement occurs

across a set of pairwise relations, the relations are coherent.

Coherence is calculated by first computing a derived score for each pair of terms. This derived score is based on some measure of agreement between the two terms' patterns of relations to the remaining terms.

The notion behind an inferred distance (hence a derived score) is the possibility of predicting the existence of an edge between two nodes based on the other edges in the graph. The inferred distance is computed between two nodes by examining the similarity of the two nodes' connections to other nodes. A

node's connections to other nodes is called its neighborhood set. In the case of graphs, a neighbor is defined to be any node that is within some path distance, exclusive of the node itself. Typically, a path distance of one edge is used to define neighbors. The Jaccard index is then used to compute the similarity of the two neighborhood sets. The Jaccard index is the ratio of the cardinality of the intersection of the sets to the cardinality of the union of the sets. Neither of the two nodes themselves is allowed to occur in the intersection set or union set. The Jaccard index ranges from zero to one. The inferred graph distance is one minus the Jaccard value.

Once the derived scores have been calculated for each pair, the correlation between derived scores and original scores is computed. This correlations is the coherence. In other words, coherence is the correlation between the original graph distances and the inferred distances.

The coherence of each subject's set of proximity scores was computed two ways. The first method employed a 25th percentile cutoff score to determine which scores were related and which were unrelated. This is exactly the procedure used above to compute scope. This coherence is called Coh25. The second

method computed coherence separately using, in turn, all of the similarity values as cutoffs to compute the derived scores. A subject's coherence was the maximum of this set of coherences. This coherence is called CohMax. Table 5 gives both of these coherence values for each subject.

Table 5. Coherence scores for each subject.

Subject	Coh25	CohMax
1	0.144	0.633
2	0.396	0.712
3	0.352	0.372
4	0.505	0.559
5	0.125	0.350
6	0.514	0.590
7	0.385	0.440
8	0.515	0.515
9	0.409	0.535
10	0.431	0.453
11	0.643	0.643
12	0.323	0.423
13	0.477	0.581
14	0.601	0.611
15	0.605	0.605
16	0.422	0.531
17	0.442	0.584
18	0.419	0.510
19	0.428	0.648
20	0.501	0.501
21	0.449	0.495
22	0.588	0.626

Pathfinder Analyses

Pathfinder network were computed for each subject and also for the mean ratings of the entire 22 person sample. The networks were generated with Pathfinder parameters r=inf and q=n-1. This r parameter value makes the fewest assumptions about data type and is the value typically used for scaling

data. Setting q=n-1 gives the leanest network.

The Pathfinder network for the average is shown in Figure 1. Note the important role played by DYNAMIC THINKING. Its centrality in the network corresponds with its high scope. This particular network is quite lean with only 25 links present. A denser network could be generated by decreasing the value of the q parameter.

Inter-subject agreement (the focus concern of this project) was examined to determine how well the 22 subjects agreed with one another. Simple Pearson correlations of the raw ratings of each subject with every other subject were obtained. Table 6 gives the resulting correlations which range from -0.02 (between subjects 3 and 5) and 0.64 (between subjects 1 and 2). The mean across all 231 comparisons was 0.326.

Table 6. Inter-subject correlations based on the raw similarity rating values.

Subject	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	64	17	21	13	47	40	29	27	17	36	37	43	41	26	47	33	19	38	24	38	27
2		37	15	06	47	39	19	26	17	41	28	44	41	30	45	35	22	35	18	39	24
3			12	-02	28	31	17	30	16	35	20	20	26	42	32	30	26	37	20	28	35
4				40	20	28	50	32	53	29	45	29	51	28	15	51	28	28	49	31	36
5					14	22	26	06	23	09	26	30	20	07	10	21	05	13	27	18	13
6						39	11	31	18	46	34	40	42	35	39	37	23	40	19	51	23
7							20	25	27	44	37	45	36	31	37	34	20	37	34	39	38
8								28	46	32	26	16	43	27	20	42	36	18	49	19	40
9									28	36	26	32	40	43	28	36	21	39	35	32	44
10										33	27	26	37	28	16	45	32	22	53	23	51
11											37	26	41	56	44	58	46	39	46	43	57
12												40	48	24	32	39	28	32	35	41	44
13													41	25	44	35	21	37	24	38	26
14														36	34	47	32	40	43	50	45
15															39	54	35	44	35	29	47
16																34	25	34	29	40	41
17																	37	31	46	36	61
18																		30	32	29	41
19																			25	45	35
20																				24	51
21																					36

The agreement of the subjects based on their individual Pathfinder networks was computed. This was accomplished by correlating corresponding graph distances from pairs of the networks. The correlations are shown in Table 7.

The correlations range from -0.10 to 0.55 with a mean of 0.20.

Table 7. Inter-subject correlations based on the 25th percentile transformed ratings.

Subject	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	38	21	19	15	14	13	38	-04	17	19	12	18	34	08	-04	11	21	31	23	33	31
2		29	09	14	29	17	13	12	27	27	07	37	31	04	-10	26	13	22	27	39	35
3			-02	04	05	-07	03	13	16	19	13	13	15	24	04	22	08	23	06	21	25
4				18	31	28	28	03	43	36	34	28	21	09	27	47	16	24	21	39	35
5					12	27	01	12	16	-02	20	20	20	03	02	23	-07	18	-02	09	09
6						09	18	16	37	25	23	26	20	-05	17	33	03	-02	14	10	32
7							10	21	17	18	14	42	25	05	03	20	09	32	15	14	15
8								03	28	43	19	01	18	20	26	19	41	28	34	16	39
9									-03	23	00	12	09	16	-01	16	00	09	10	10	14
10										28	40	27	24	03	34	44	27	29	24	18	47
11											25	20	22	38	18	47	35	30	36	43	46
12												20	40	02	30	41	11	27	07	20	22
13													29	09	12	28	03	38	11	31	19
14														09	03	28	08	34	-06	32	22
15															14	33	13	50	11	21	34
16																18	23	15	12	15	21
17																	09	37	19	34	55
18																		24	40	28	34
19																			27	35	45
20																				35	42
21																					41

A check was made to explore for especially low agreements among the expert subjects. The data of Table 6 was put into the following rectangular matrix and each row of correlations summed and divided by 21.

Table 8 Correlations from Table 6 in rectangular form

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22
1	0	64	17	21	13	47	40	29	27	17	36	37	43	41	26	47	33	19	38	24	38	27
2	64	0	37	15	6	47	39	19	26	17	41	28	44	41	30	45	35	22	35	18	39	24
3	17	37	0	12	-2	28	31	17	30	16	35	20	20	26	42	32	30	26	37	20	28	35
4	21	15	12	0	40	20	28	50	32	53	29	45	29	51	28	15	51	28	28	49	31	36
5	13	6	-2	40	0	14	22	26	6	23	9	26	30	20	7	10	21	5	13	27	18	13
6	47	47	28	20	14	0	39	11	31	18	46	34	40	42	35	39	37	23	40	19	51	23
7	40	39	31	28	22	39	0	20	25	27	44	37	45	36	31	37	34	20	37	34	39	38
8	29	19	17	50	26	11	20	0	28	46	32	26	16	43	27	20	42	36	18	49	19	40
9	27	26	30	32	6	31	25	28	0	28	36	26	32	40	43	28	36	21	39	35	32	44
10	17	17	16	53	23	18	27	46	28	0	33	27	26	37	28	16	45	32	22	53	23	51
11	36	41	35	29	9	46	44	32	36	33	0	37	26	41	56	44	58	46	39	46	43	57
12	37	28	20	45	26	34	37	26	26	27	37	0	40	48	24	32	39	28	32	35	41	44
13	43	44	20	29	30	40	45	16	32	26	26	40	0	41	25	44	35	21	37	24	38	26
14	41	41	26	51	20	42	36	43	40	37	41	48	41	0	36	34	47	32	40	43	50	45
15	26	30	42	28	7	35	31	27	43	28	56	24	25	36	0	39	54	35	44	35	29	47
16	47	45	32	15	10	39	37	20	28	16	44	32	44	34	39	0	34	25	34	29	40	41
17	33	35	30	51	21	37	34	42	36	45	58	39	35	47	54	34	0	37	31	46	36	61
18	19	22	26	28	5	23	20	36	21	32	46	28	21	32	35	25	37	0	30	32	29	41
19	38	35	37	28	13	40	37	18	39	22	39	32	37	40	44	34	31	30	0	25	45	35
20	24	18	20	49	27	19	34	49	35	53	46	35	24	43	35	29	46	32	25	0	24	51
21	38	39	28	31	18	51	39	19	32	23	43	41	38	50	29	40	36	29	45	24	0	36
22	27	24	35	36	13	23	38	40	44	51	57	44	26	45	47	41	61	41	35	51	36	0

Table 9: Mean Correlations by Subject from Table 8 Data

Subject #	Mean Score
1	32.57
2	32.00
3	25.57
4	32.90
5	16.52*
6	32.57
7	33.48
8	29.24
9	30.71
10	30.38
11	39.71
12	33.62
13	32.48
14	39.71
15	34.33
16	32.62
17	40.10
18	28.00
19	33.29
20	34.19
21	34.71
22	38.81

*≈Subject 5 stands out.

The mean agreement among the US sample and the remaining subjects was examined. The mean correlation among all possible comparisons (28) of the 8 US subjects was 0.3732. The mean correlation among all possible

comparisons (91) of the remaining 14 subjects was 0.3057.
The mean and standard deviation for each pair of terms across the 22 subjects was computed and is provided in the Appendix. Finally, the mean standard deviation for each term was calculated by averaging the standard deviations of those pairs of terms that contained the term in question. These values are reported in Table 10.

This information identifies which of the terms the experts tended to agree on (i.e., showed low variability in their ratings with other terms) and which of the terms produced the most disagreement. The raw ratings by pair for the 22 subjects are

are provided in the Appendix along with mean and standard deviation.

Table 10. Mean standard deviation of each term ranked from least to most deviation.

Term	Mean Std
DYNAMIC THINKING	1.88
SENSITIVITY TESTING	1.89
ROBUSTNESS	1.92
SIGMOIDAL GROWTH	1.94
OSCILLATION	1.96
LEVEL (STOCK) VARIABLE	1.97
NEGATIVE FEEDBACK	1.97
COMPENSATING FEEDBACK	1.99
CONTROL STRUCTURE	2.01
PARAMETER INSENSITIVITY	2.01
DYNAMIC HYPOTHESIS	2.03
ENDOGENOUS VIEW	2.04
POSITIVE FEEDBACK	2.05
EXTREME CONDITIONS TEST	2.05
ACCUMULATION	2.06
OVERSHOOT AND COLLAPSE	2.06
EXPONENTIAL GROWTH	2.07
DOMINANCE SHIFTS	2.07
AGGREGATION	2.09
BOUNDARY	2.12
CAUSATION	2.15
RATE (FLOW) VARIABLE	2.15
NONLINEAR RELATIONSHIP	2.17
EXPONENTIAL DECAY	2.34
OPERATIONAL THINKING	2.35

A comparison of the mean agreement among US vs non-US subject was done. The mean correlation among all possible comparisons (28) of the 8
US subjects was 0.3732. The mean correlation among all possible comparisons (91) of the remaining 14 subjects was 0.3057.

INTERPRETATION

The principal finding of this research is that the similarity scores indicative of agreement among expert system dynamists' conceptual structures is somewhat lower than like conceptual structure comparisons among experts in other fields (the usual similarity scores range from .4 to .6). There are several possible reasons for the results obtained. First, many of the subjects may not have achieved "expert" level knowledge of system dynamics despite the selection criteria and methods used to identify people with expertise. The data may be pointing out flaws in the assumption that advanced degrees and experience in the field are reliable indicator of expert achievement. In other words, the sample may not have been as homogeneous vis-a-vis system dynamics expertise as was thought.

Second, it is possible that the concept set (in part or in entirety) presented to the subjects was not recognized, or accepted, by the

subjects as representing the core concepts of the field. This is germane as the instructions specifically asked the subjects to make relatedness judgments in terms of core concepts of system dynamics.

A third possibility is that there is not wide agreement within the field about the meaning of the terms in the concept set and/or how such terms are related. Put a different way, this data may indicate that the field of system dynamics is in a formative stage.

Finally, the subjects may indeed all be expert level workers in the field, but due to their very heterogeneous cultural, national, and educational backgrounds may understand and interpret the similarities of the concepts differently. Virtually all of the research on expertise using this approach has employed a much more uniform subject group. The influence of gender is unknown because only one female agreed to participate as a subject.

While the first two possibilities may have contributed to the pattern of the data, the latter two interpretations seem most plausible. System Dynamics is a relatively new field and many of its concepts have been borrowed and/or modified from other disciplines (Richardson, 1991). This point is reflected in the results reported in Table 10. Additionally,

from the limited analysis of mean agreements among US vs non-US subjects, it does seem that experience, in the most general sense, influences expert level understanding.

DISCUSSION

Several things were learned as a consequence of the research project. The PathPrep software tool performed up to all expectations. The task of selecting a subset of terms from a larger set that delimited the knowledge domain was made simpler and more efficient than other techniques such as card sorts. Additionally, not one of the subjects reported any problems using the tool to make similarity judgments. No query of any kind was made by the subjects requiring assistance in the operation of the PathPrep judgment tool. Every datum from each subject was present in every diskette received. The point is that the strategy of data collection via magnetic diskette is feasible, at least if one is working with knowledgeable subjects and well designed software.

It was expected at the beginning of the project that the results would show a fairly robust similarity score among the experts. The fact that weaker scores were obtained was a disappointment. It would be interesting to do the same experiment with comparable subjects in a more established field like physics or biology to ascertain the influence

of the knowledge domain per se on subjects' judgments. Given the lack of research on this question, a lesson from this study is to take a conservative approach to sample selection and investigate expert conceptual structure properties among a more homogeneous group.

This research was presented in a more general context having to do with measuring changes in knowledge domains which change the way students view the world. The idea was to see if a general standard existed which could be used as a criterion base to judge students' knowledge acquisition, particularly in a computer based interactive learning environment. The answer is a probable no and a likely yes.

It is no to the extent that, at least for system dynamics, reasonable experts do differ in the way they organize their knowledge and hence understand their field. This might explain why people can develop quite different models of problems which nonetheless achieve their goals.

The answer is a tentative yes if it is the conceptual structure of a particular teacher that is used as the standard of comparison against which student conceptual structures are assessed. Such comparisons may serve as a method to diagnose misconceptions and to determine what the next instructional message ought to be. Such

"snapshots",taken periodically, might reveal patterns of knowledge acquisition during the student's progress from novice toward expertise. This use of conceptual structure analyses is still a matter for investigation, but the early indications are promising.

SUMMARY AND CONCLUSIONS

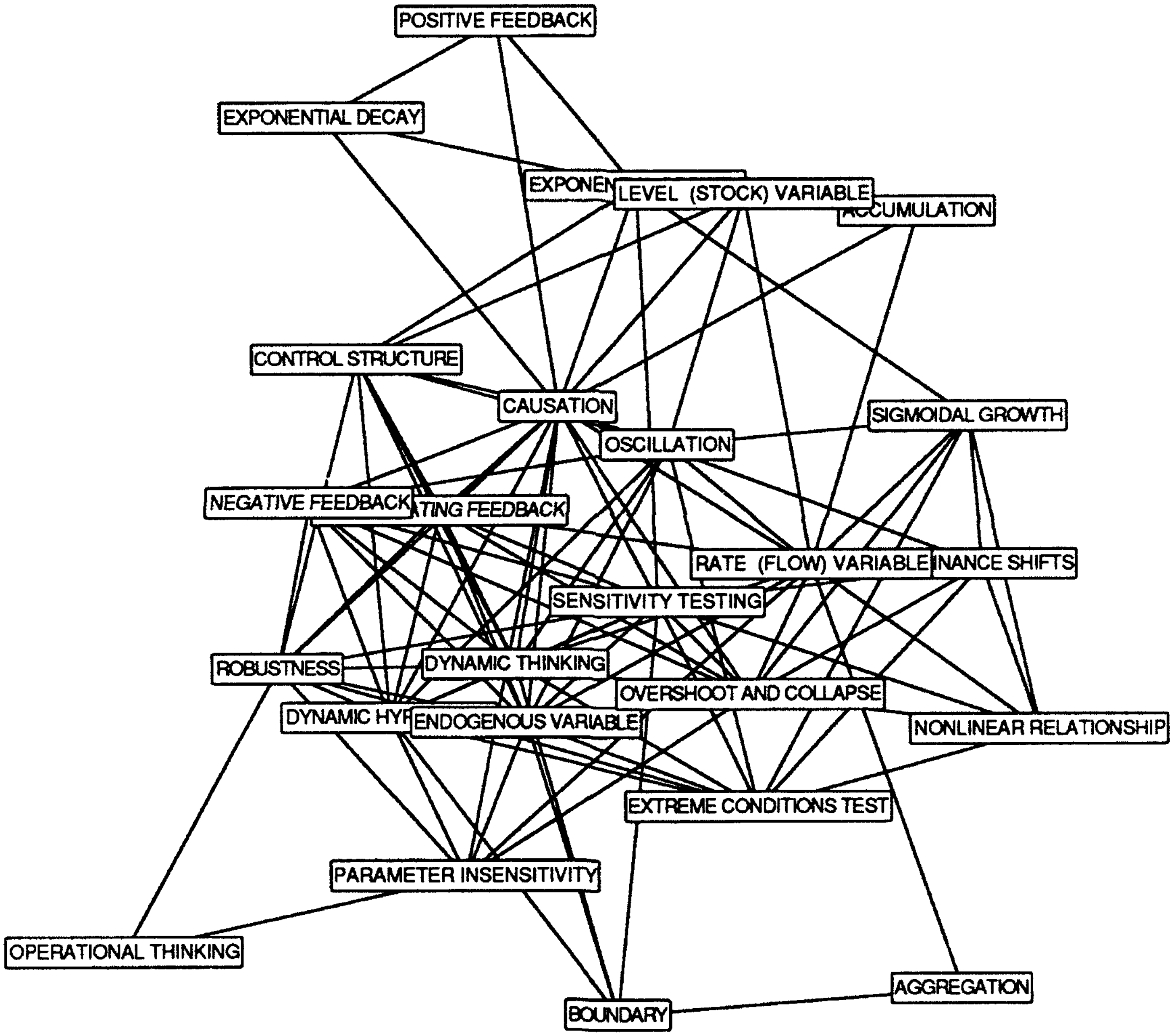
This project investigated the extent to which a diverse group of expert system dynamists had similar conceptual structures of core concepts of their discipline. The data were derived by having subjects make judgments of 25 terms presented randomly and assigning to each pair a value within the range of one to nine based on the criterion of relatedness. The data were analyzed as Pathfinder networks which emphasize structural properties of the graphs which are assumed to depict the relationships among the various concepts. The similarities found were lower than previous research on experts of other fields had achieved. Some suggestions for interpreting the findings were offered along with possible pedagogical applications.

APPENDIX

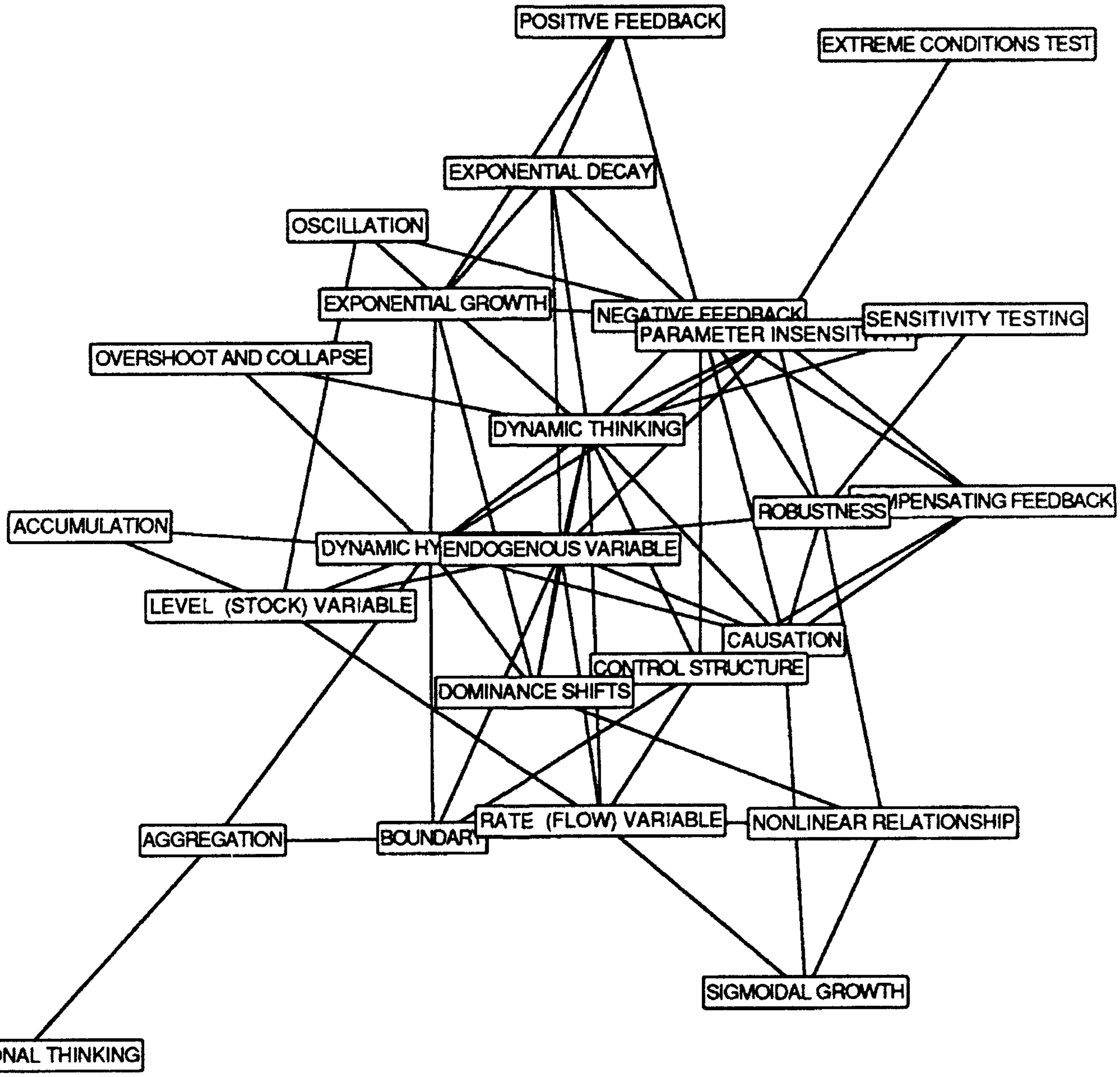
Pathfinder Graphs of Subjects' Conceptual Structures

The length of the links of the Pathfinder graphs are not accurate depictions of distances between nodes. They represent the most efficient way to display the graph on the computer monitor. However, the neighborhood depictions are meaningful and indicate significant associations among concepts. The Similarity between the graphs is presented in Tables 6 (p29) and 8 (p31)..

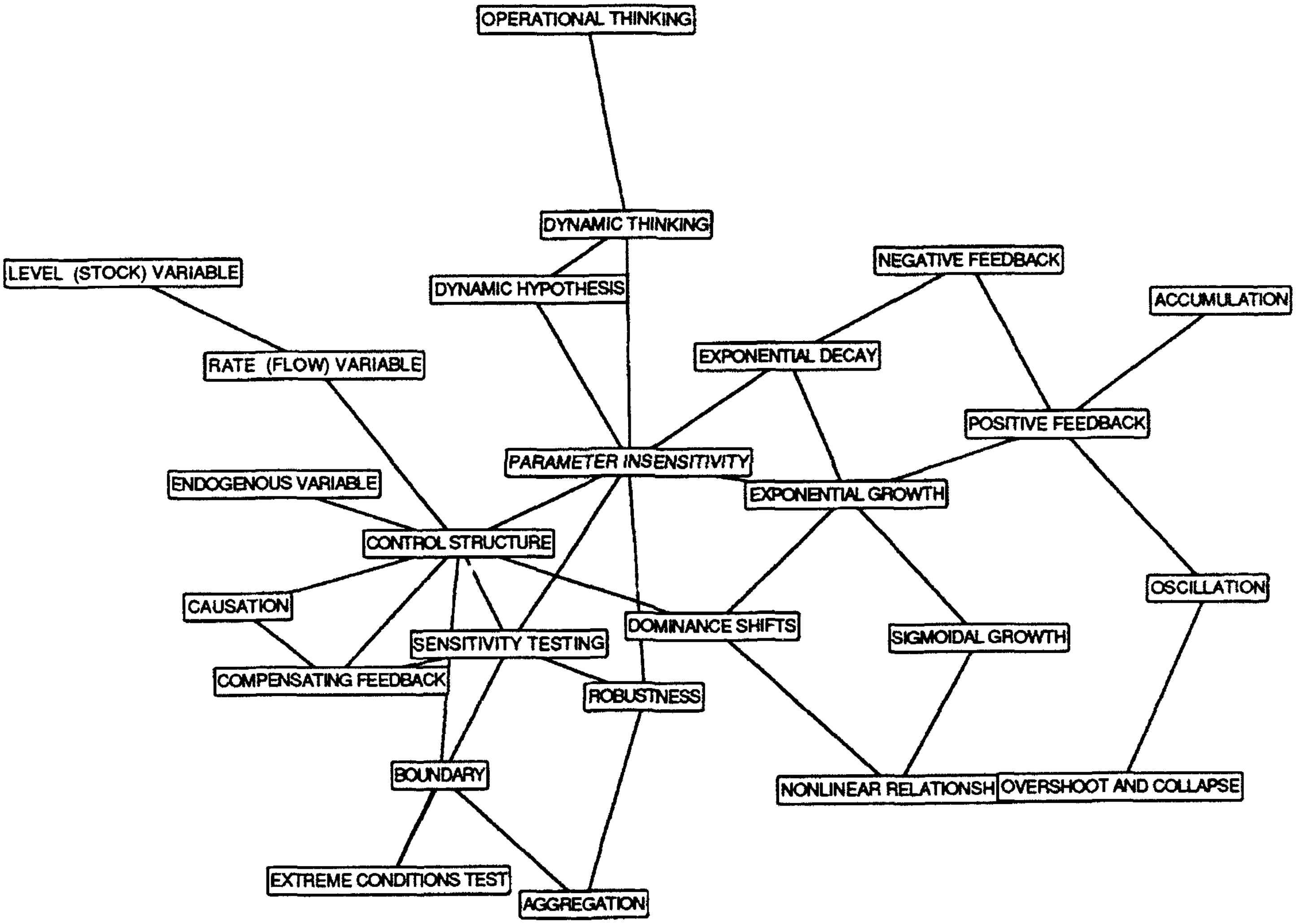
·CA89RH - q=24 - r=INF



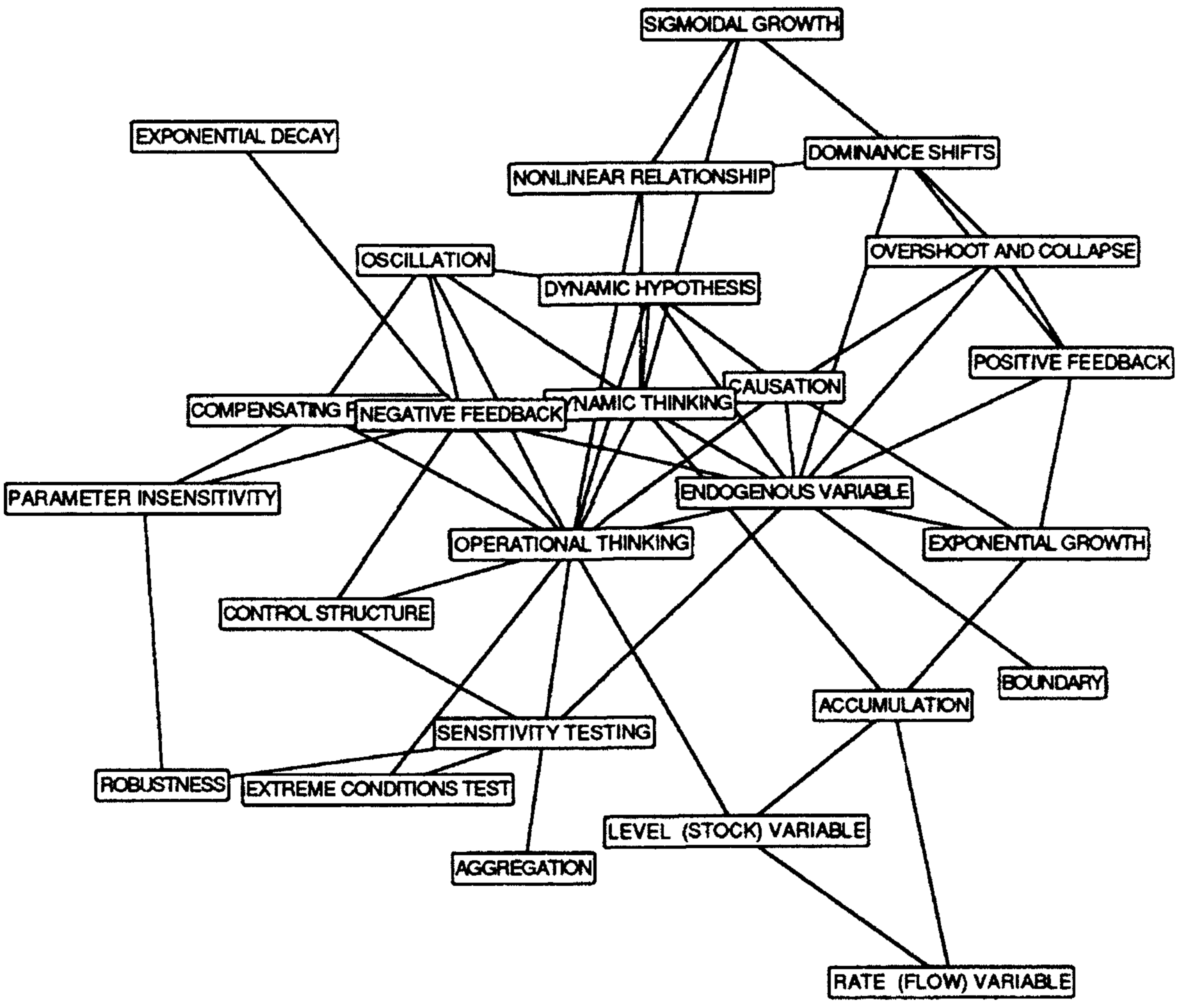
·CA20JR - q=24 - r=INF



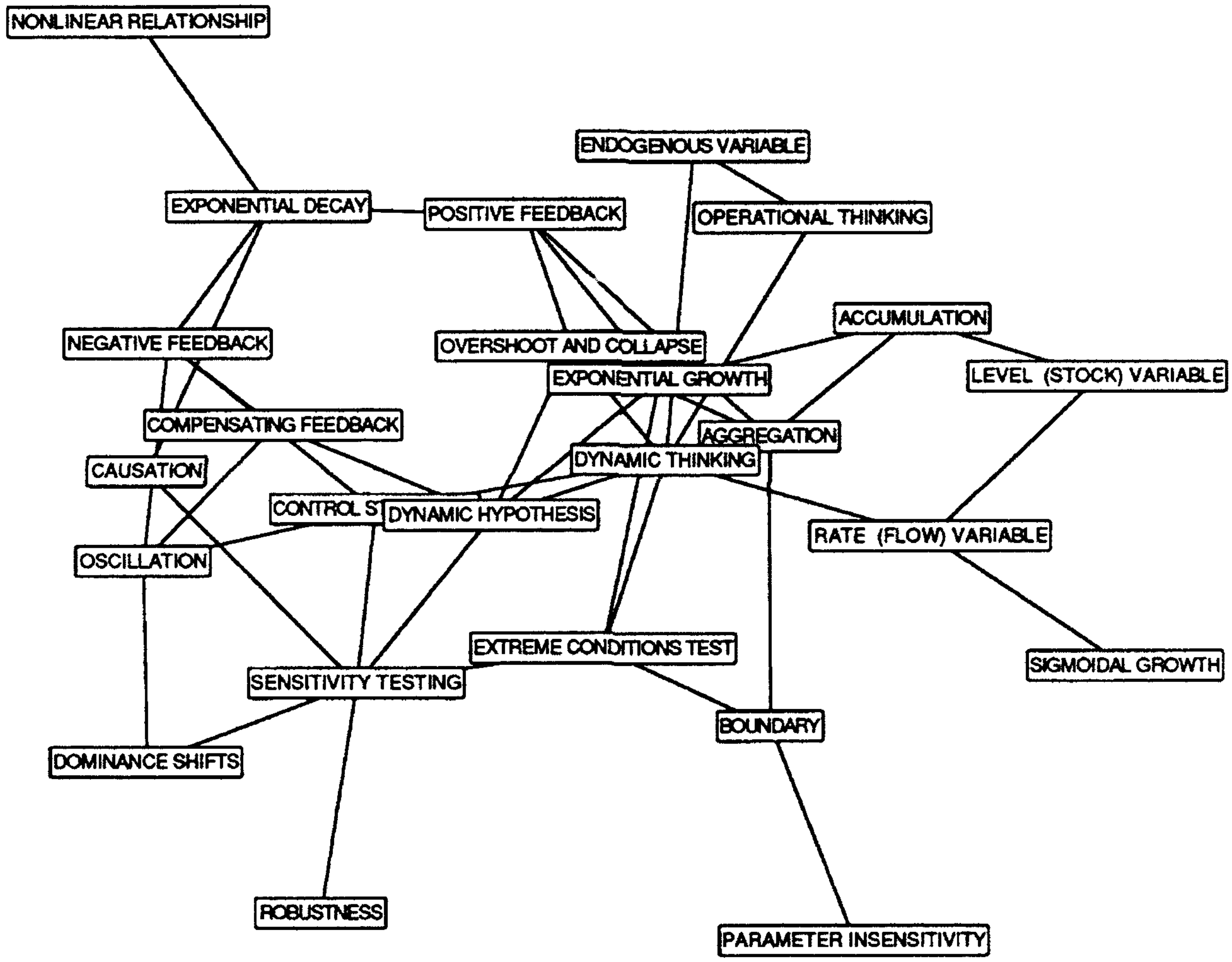
GR29PM - q=24 - r=INF



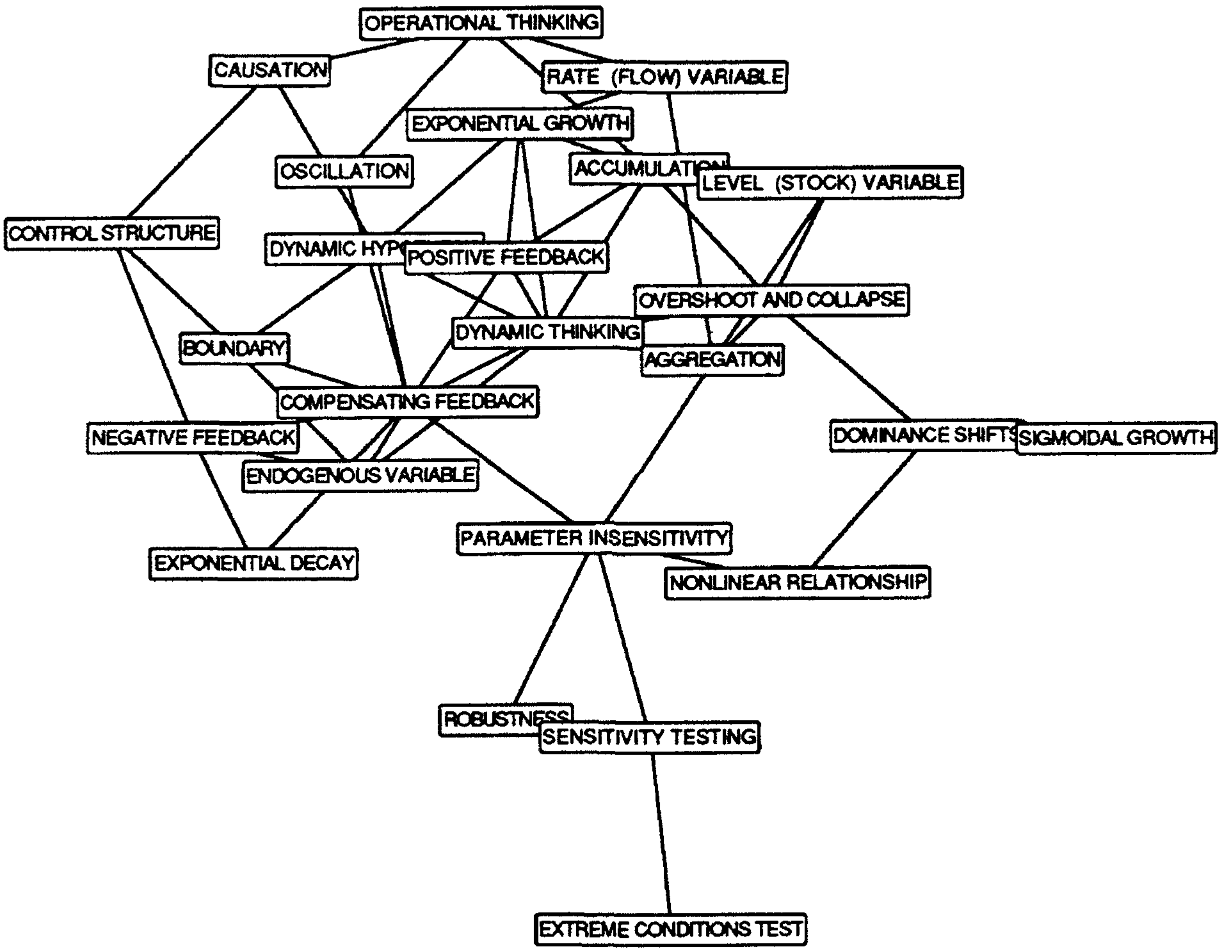
IN28PM - q=24 - r=INF



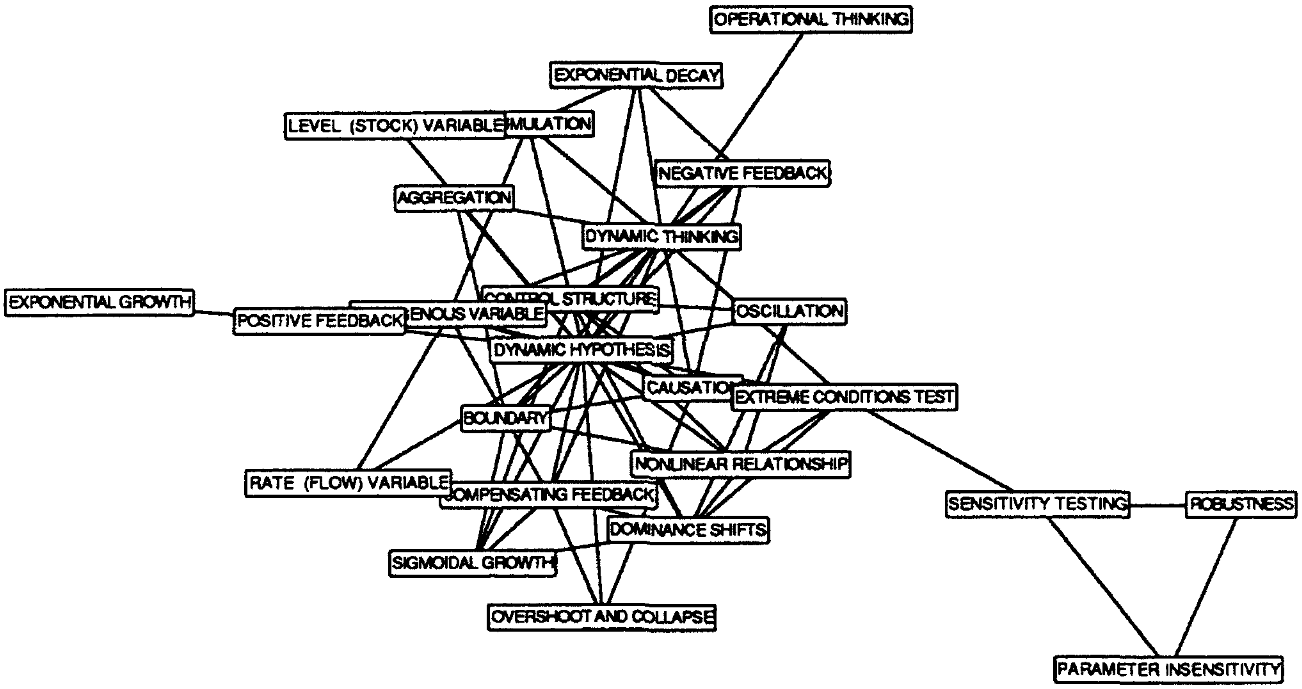
IS11CJ - q=24 - r=INF



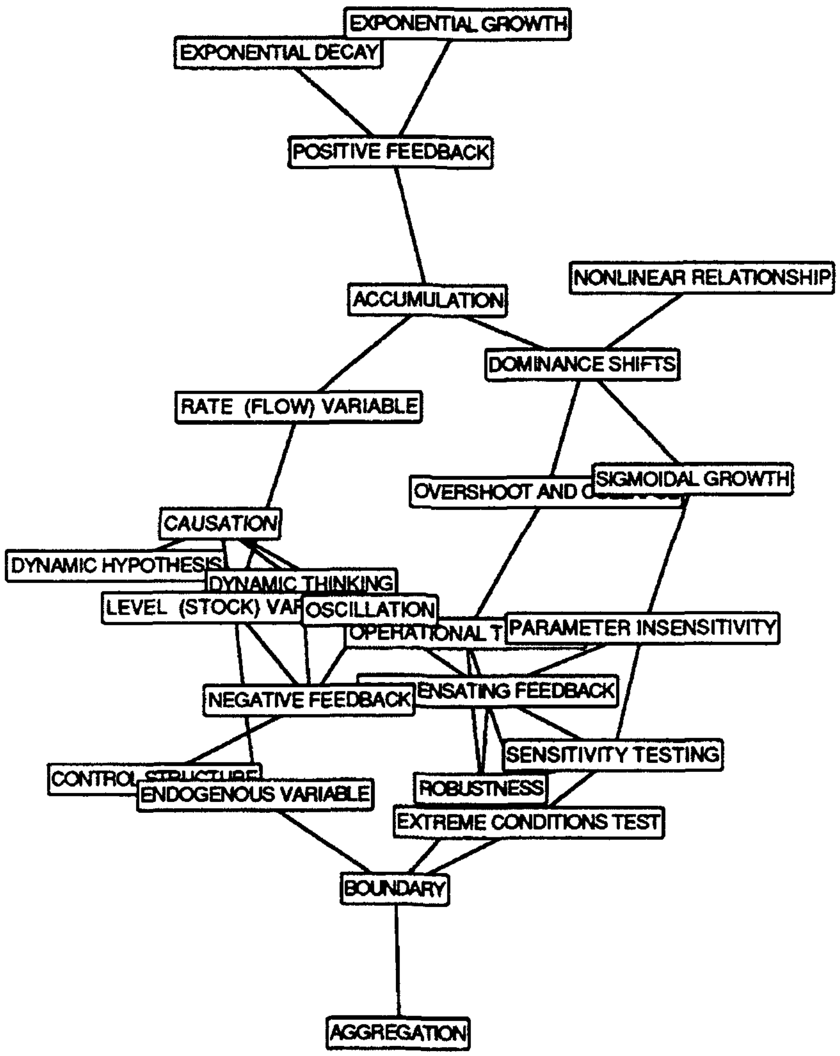
· NE26JU - q=24 - r=INF



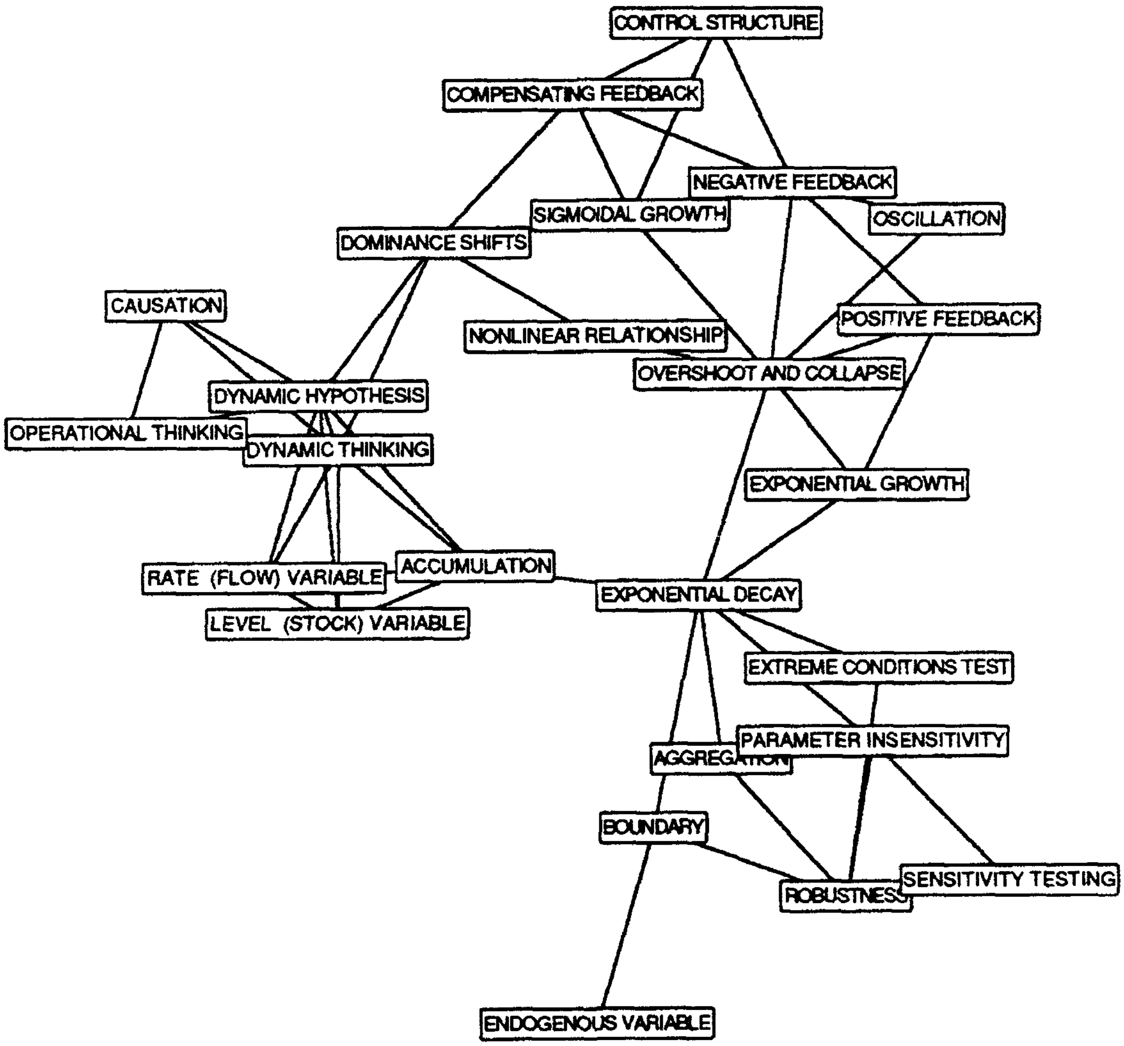
NE27WW - q=24 - r=INF



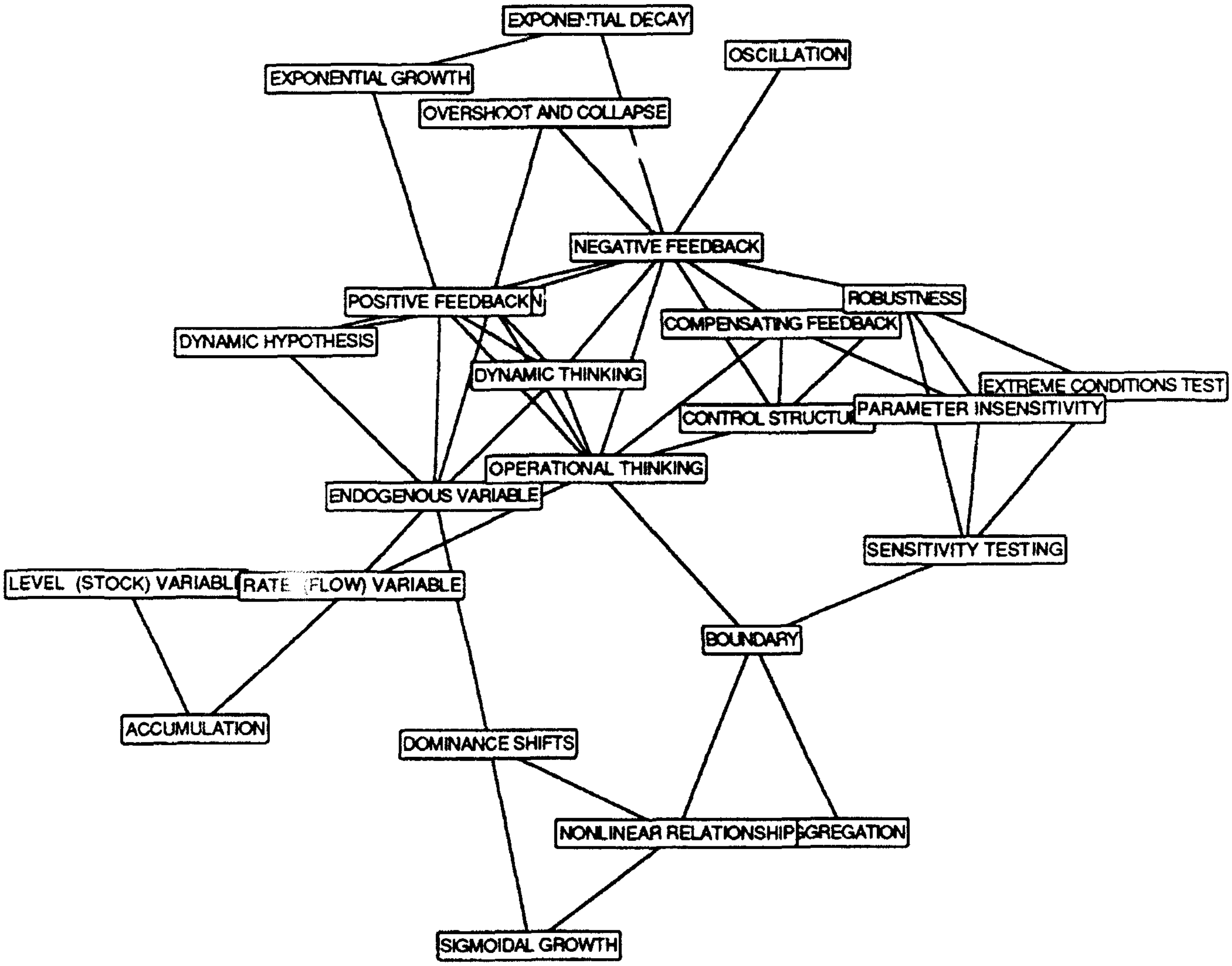
N015EM - q=24 - r=INF



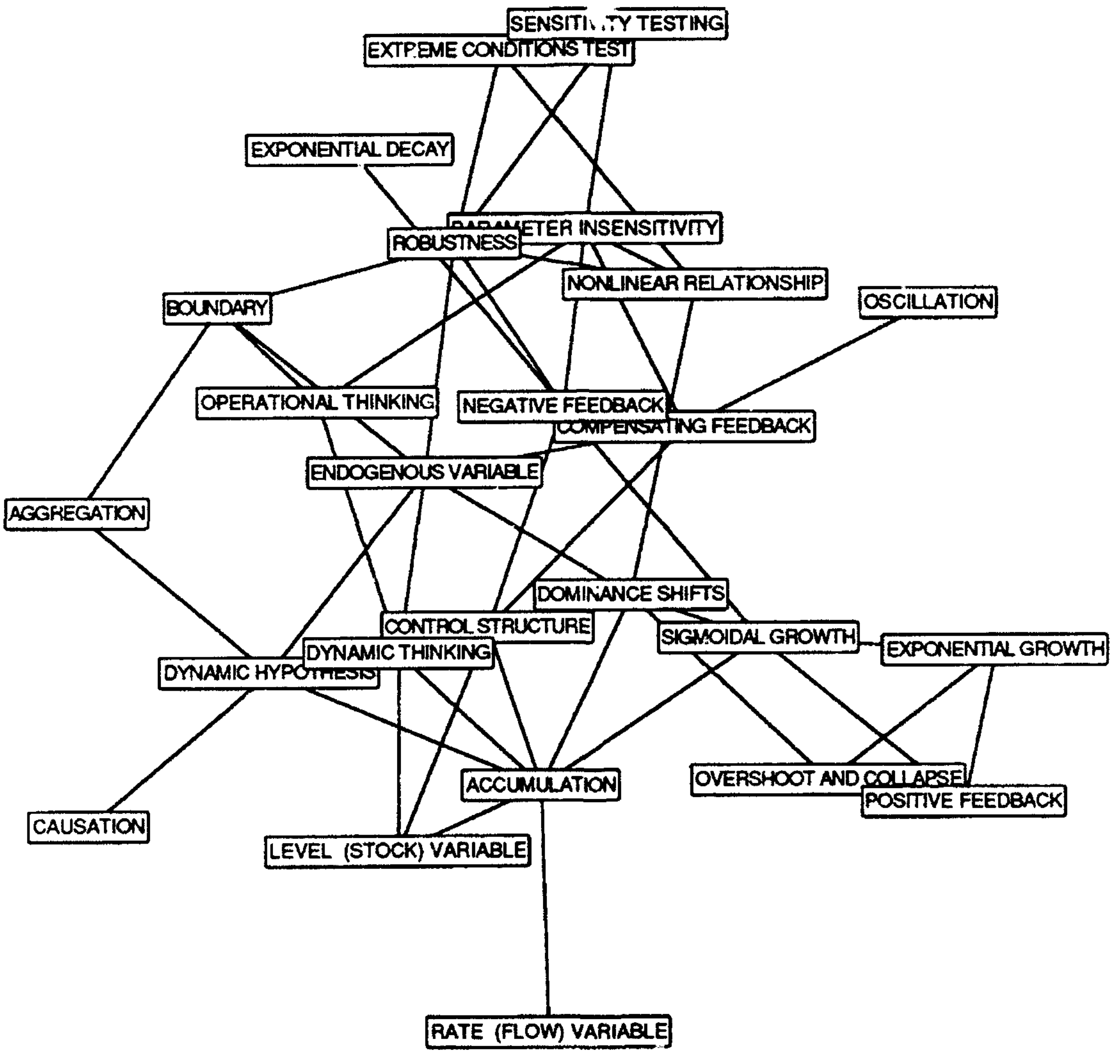
• N038PD - q=24 - r=INF



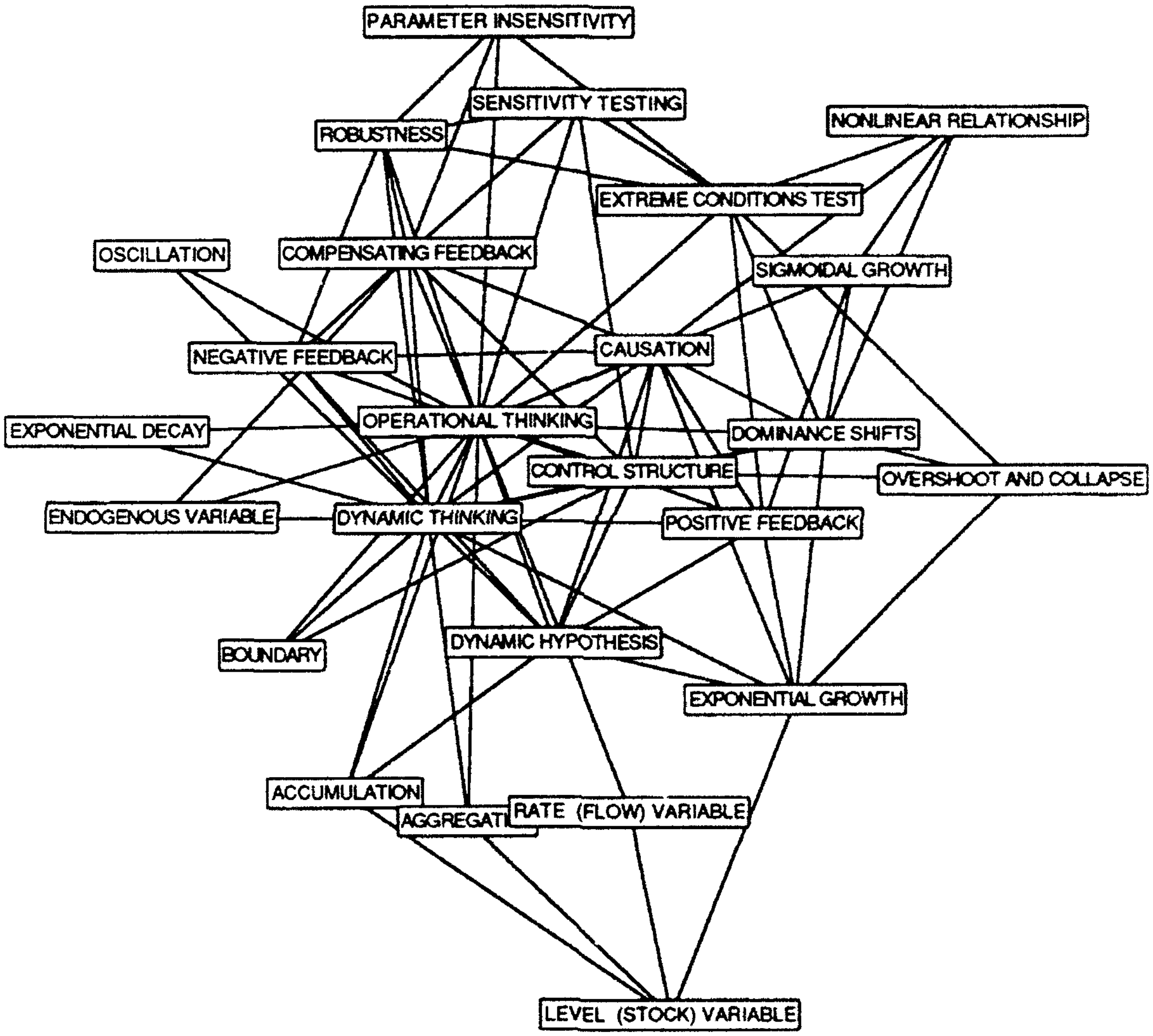
. TI19KS - q=24 - r=INF



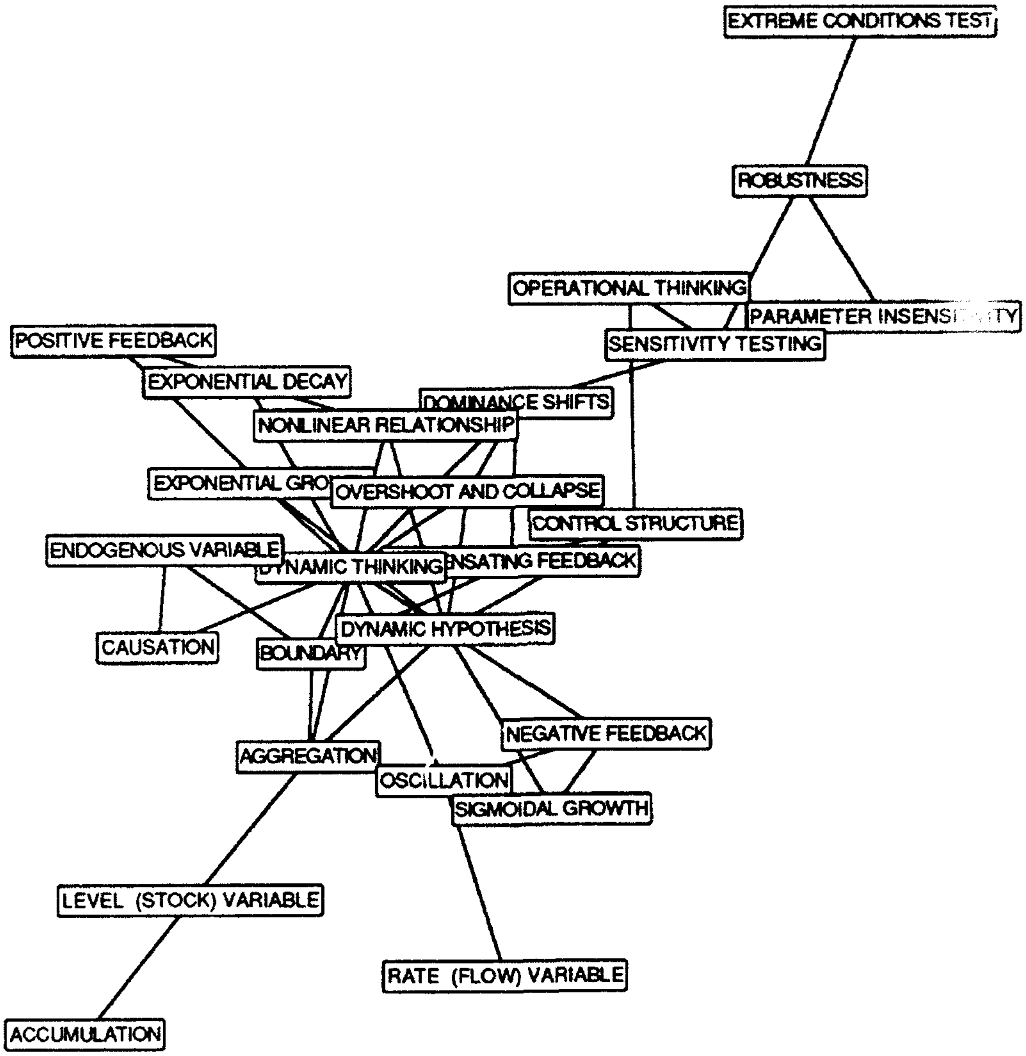
TR84YB - q=24 - r=INF



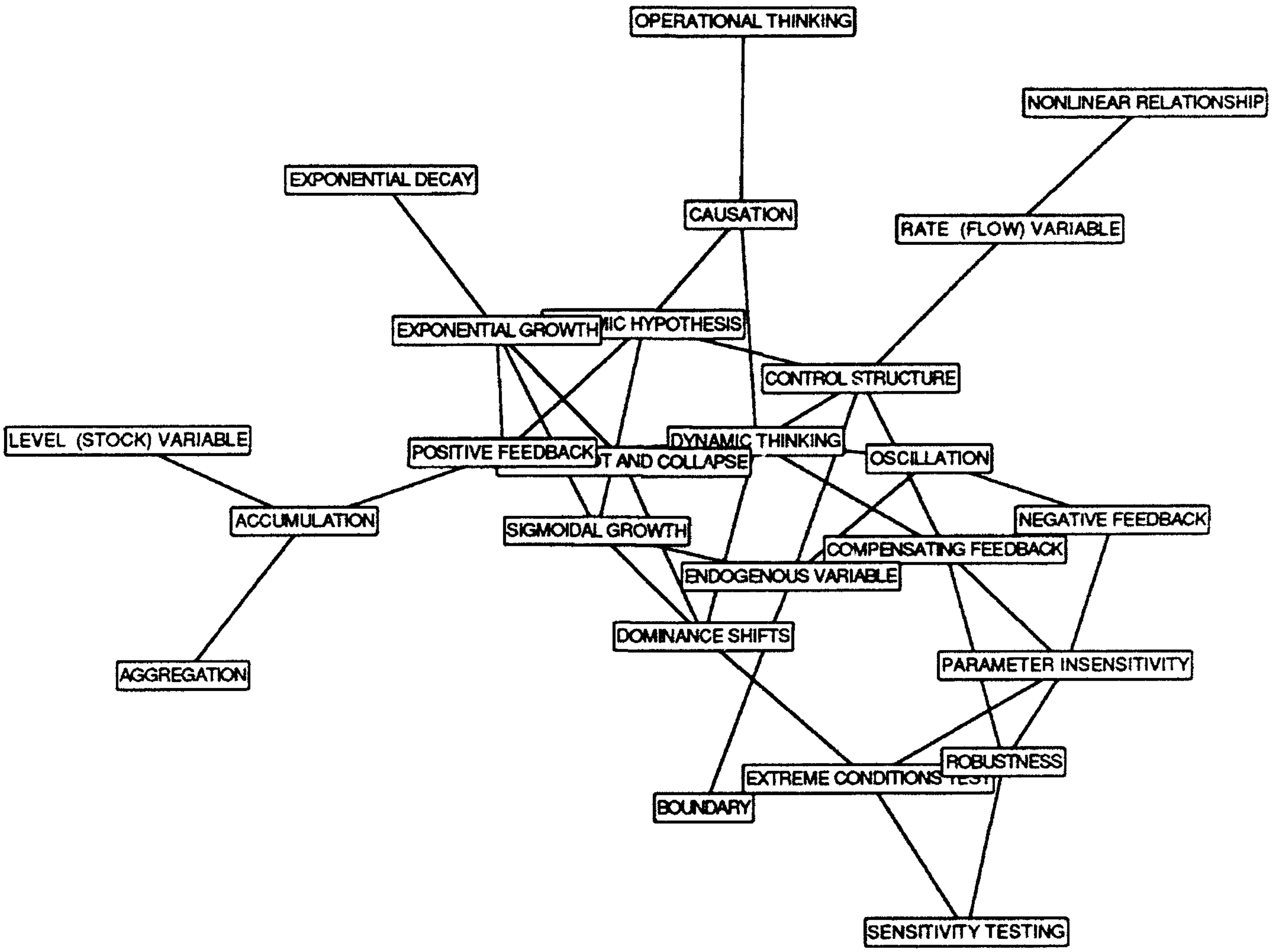
- UK05GC - q=24 - r=INF



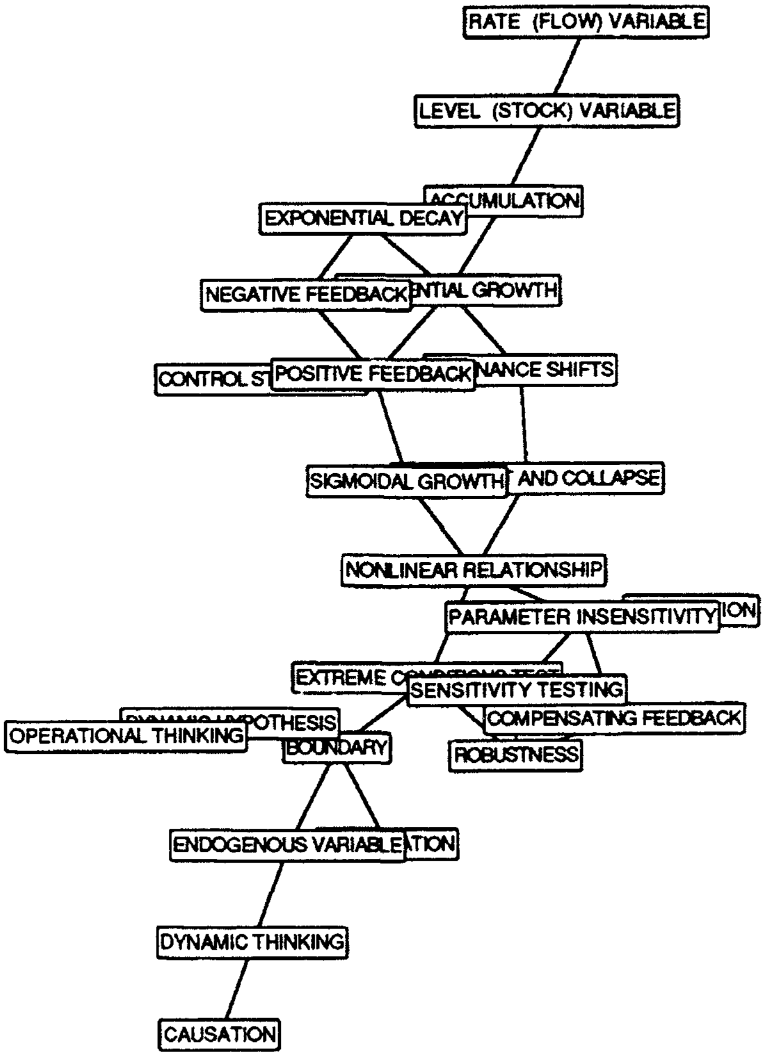
UK24EW - q=24 - r=INF



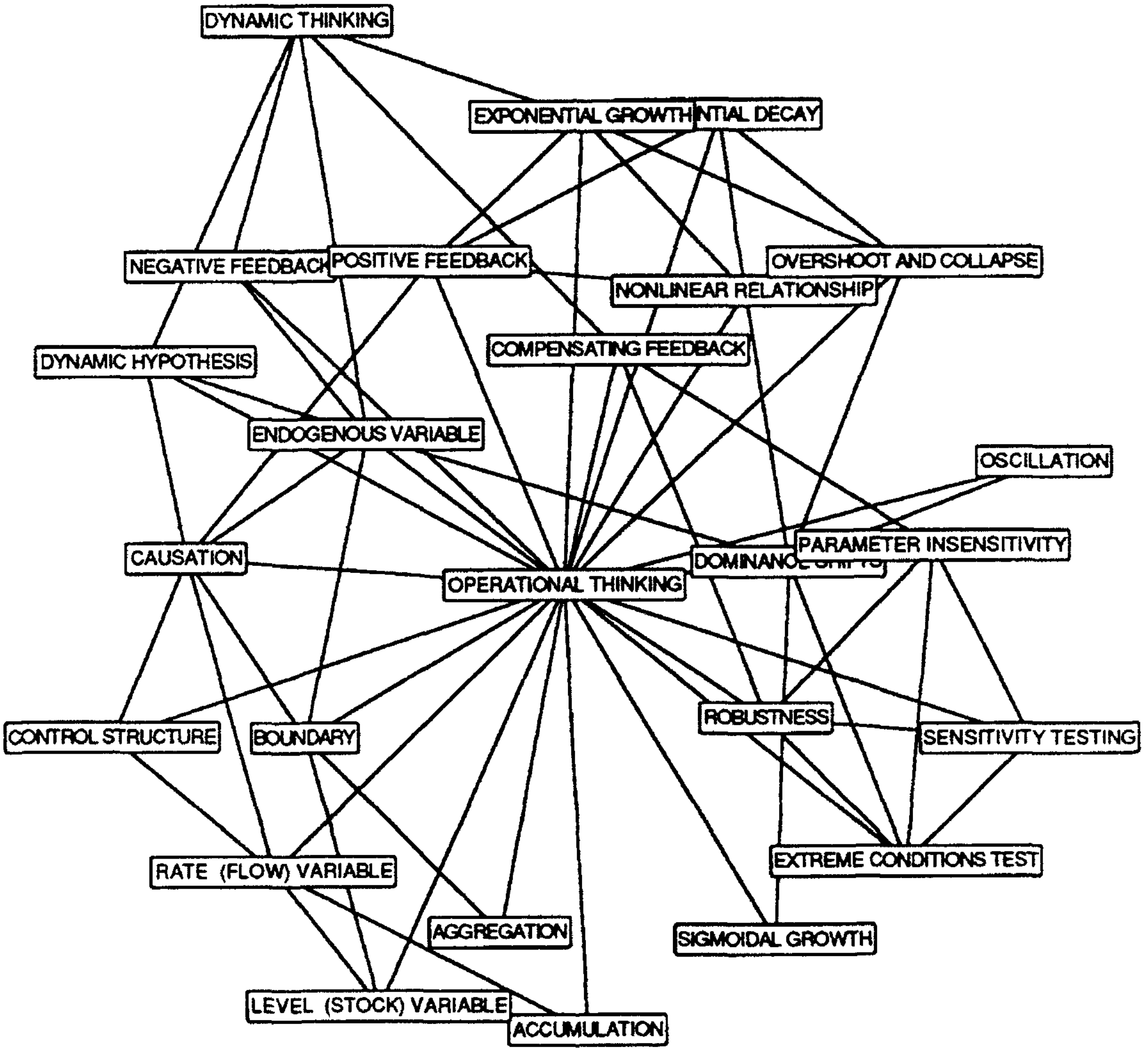
UK25GW - q=24 - r=INF



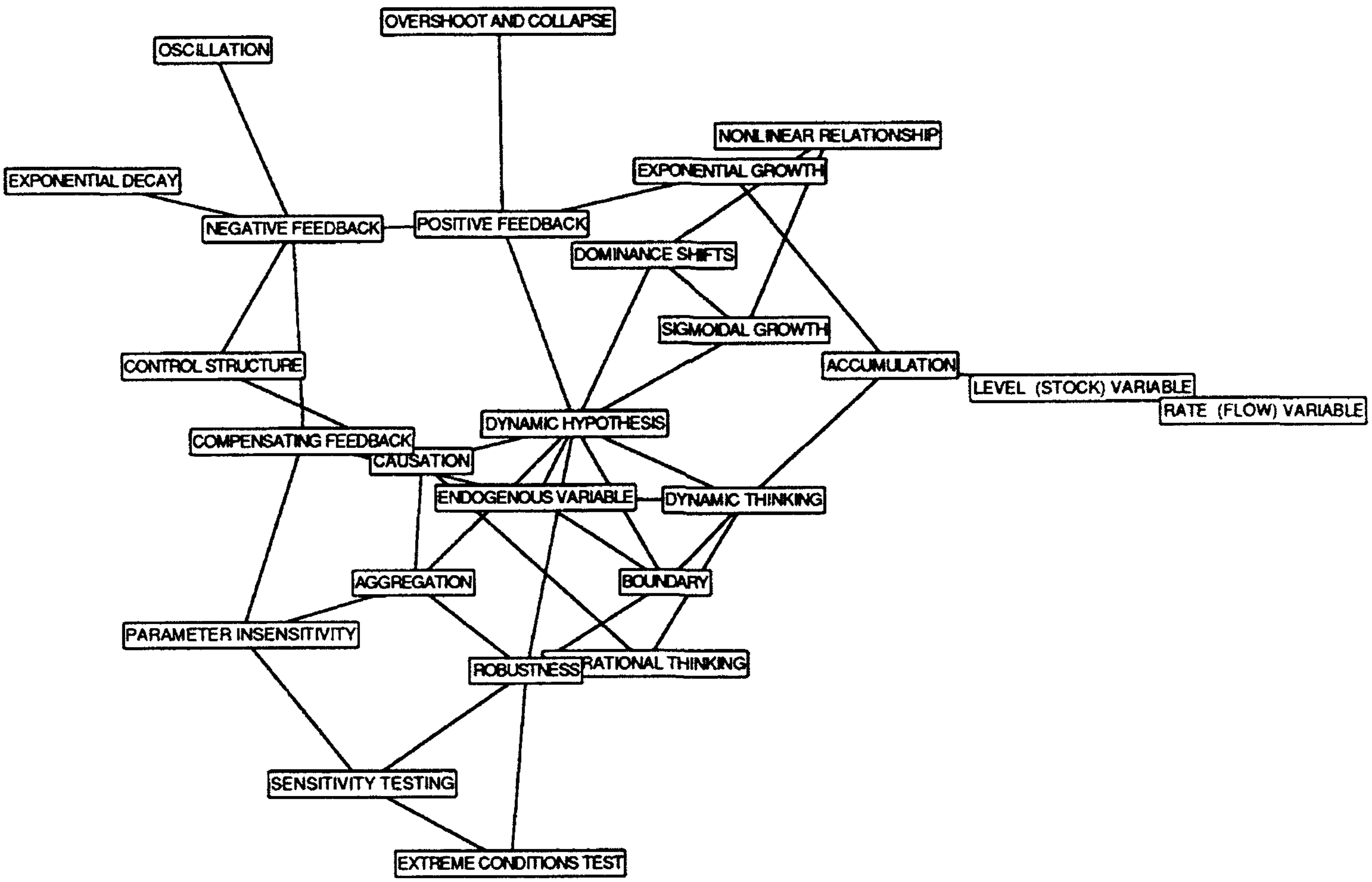
US87NF - q=24 - r=INF



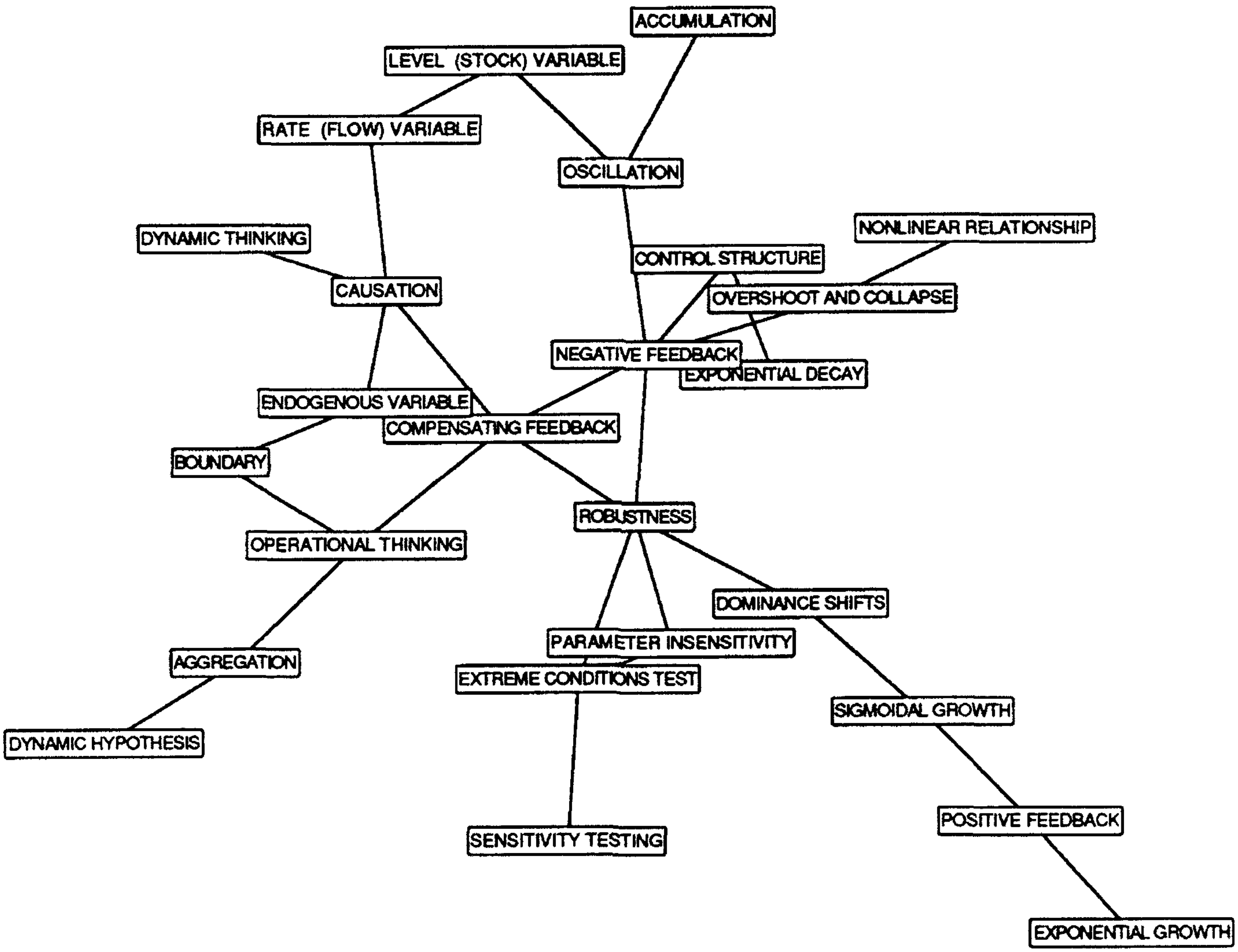
-'US88AG - q=24 - r=INF



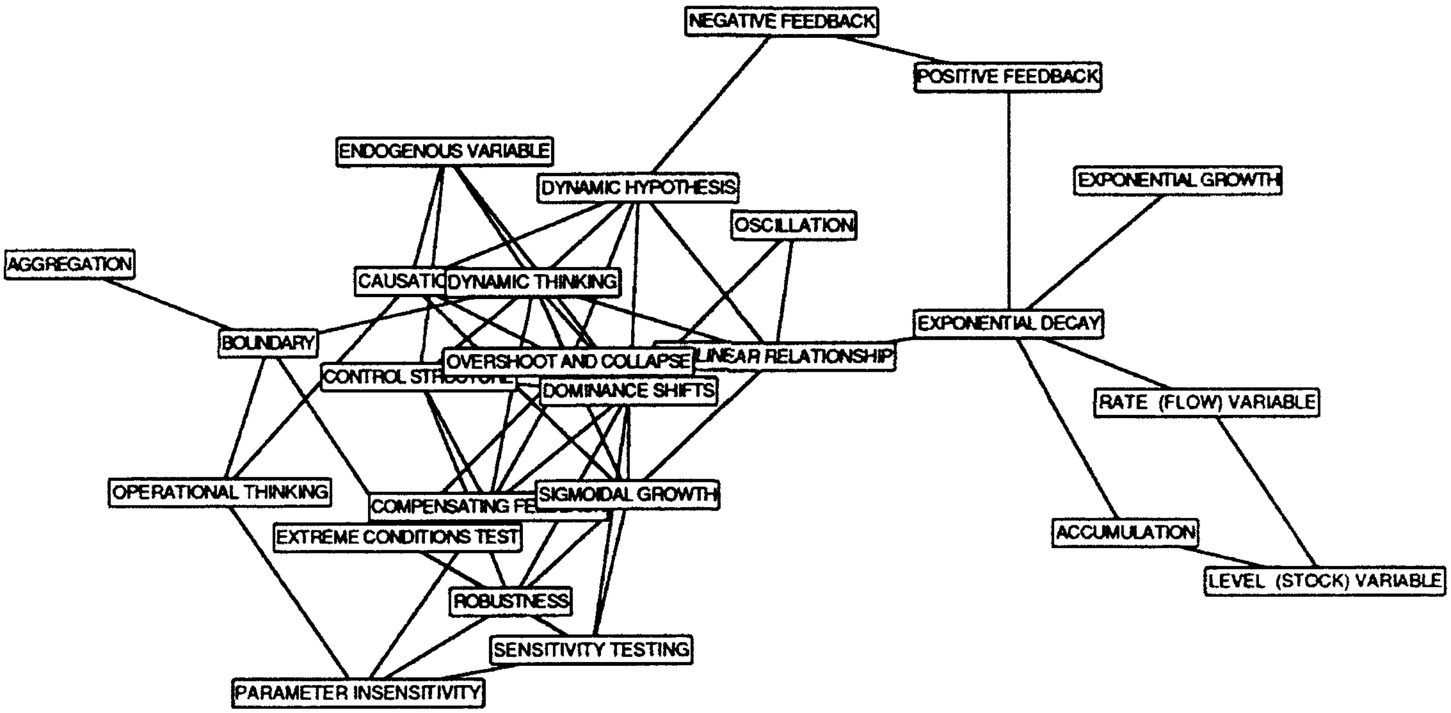
US18JR - q=24 - r=INF



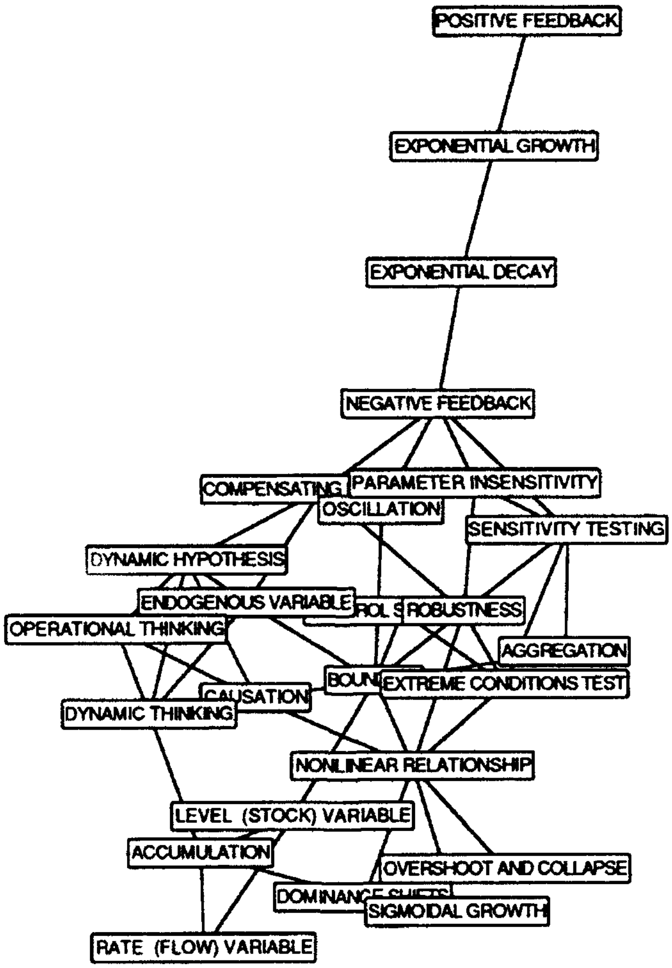
-'US14NM - q=24 - r=INF



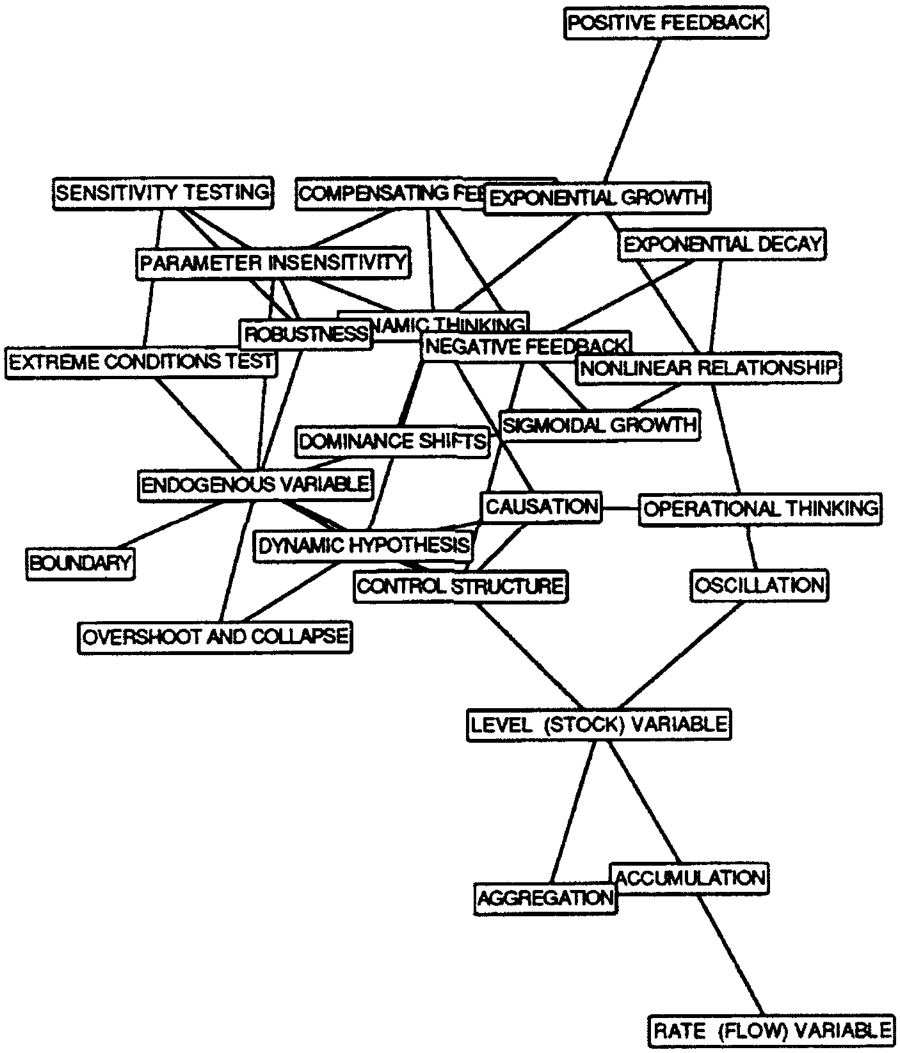
US17RQ - q=24 - r=INF



US18MR - q=24 - r=INF



US21JR - q=24 - r=INF



US23JS - q=24 - r=INF